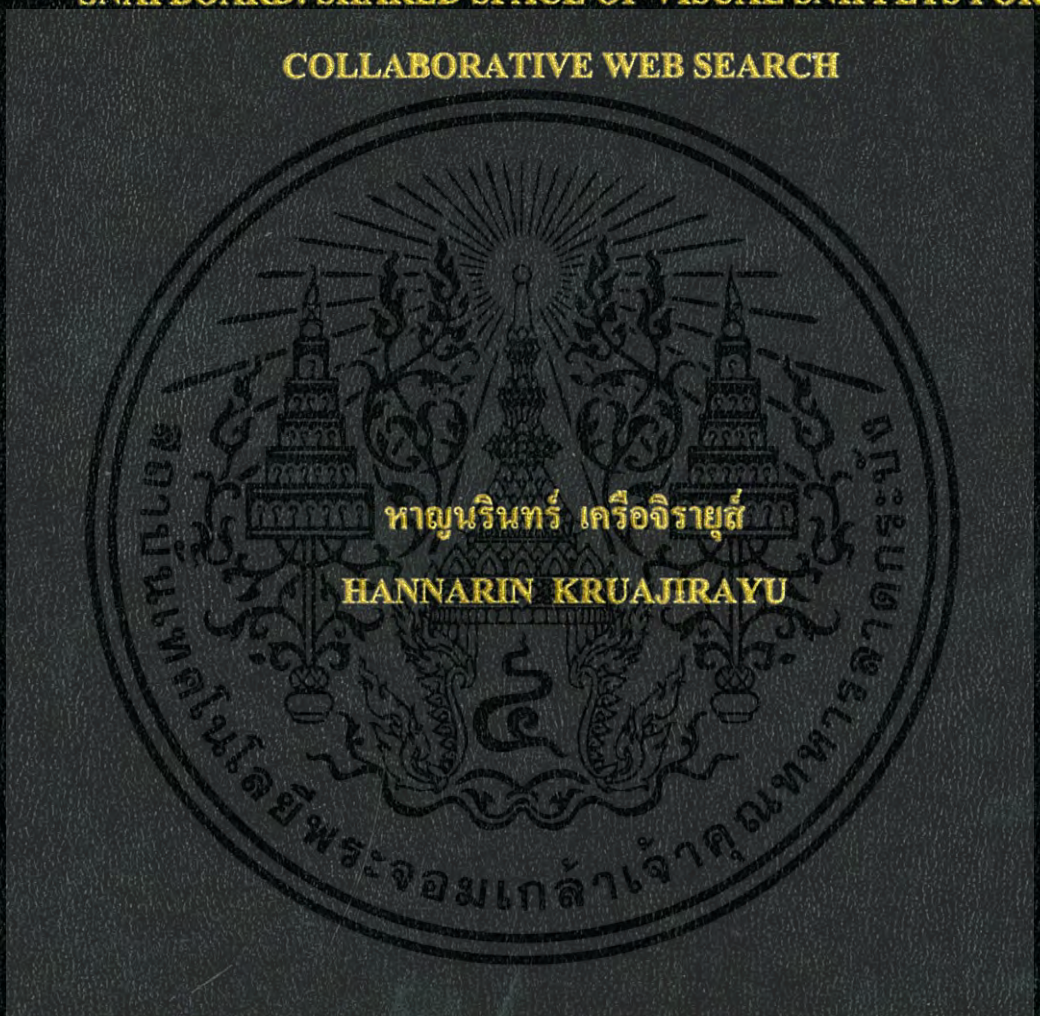
สแนปบอร์ด: กระดานแสดงและแบ่งปันสรุปแบบภาพสำหรับการค้นหาเว็บไซต์แบบร่วมมือ

SNAPBOARD: SHARED SPACE OF VISUAL SNIPPETS FOR COLLABORATIVE WEB SEARCH



หาญนรินทร์ เครือจิรายุส์

HANNARIN KRUAJIRAYU

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2558

KMITL-2015-IT-M-001-012

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

สแนปบอร์ด: กระดานแสดงและแบ่งปันสรุปแบบภาพสำหรับการค้นหาเว็บไซต์แบบร่วมมือ

SNAPBOARD: SHARED SPACE OF VISUAL SNIPPETS FOR COLLABORATIVE WEB SEARCH

T143970

หาญนรินทร์ เครือจิรายุส์

HANNARIN KRUAJIRAYU

สจหมู.....
เลขทะเบียน.....143970
วัน.เดือน.ปี.....10 ต.ค. 2559

.b.00266952
.i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2558

KMITL-2015-IT-M-001-012

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SNAPBOARD: SHARED SPACE OF VISUAL SNIPPETS FOR COLLABORATIVE WEB SEARCH

HANNARIN KRUAJIRAYU

A THESIS SUBMITTED IN FULFILLMENT OF

THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2015

KMITL-2015-IT-M-001-012

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2015

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ สแนปบอร์ด: กระดานแสดงและแบ่งปันสรุปแบบภาพสำหรับการค้นหาเว็บไซต์แบบร่วมมือ

Snapboard: Shared space of visual snippets for collaborative web search

นักศึกษา นายหาญนรินทร์ เครือจิรายุส์

รหัสประจำตัว 57606150

ปริญญา วิทยาศาสตรมหาบัณฑิต

สาขาวิชา เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.ธีรพงศ์ ลีลานุภาพ

คณะกรรมการสอบวิทยานิพนธ์	ลายมือชื่อ
รองศาสตราจารย์ ดร.อาริต ธรรมโน	
ดร.อลิสา คงทน	
ผู้ช่วยศาสตราจารย์ ดร.ธีรพงศ์ ลีลานุภาพ	
ผู้ช่วยศาสตราจารย์ ดร.กิติ์สุชาต พสุภา	
รองศาสตราจารย์ ดร.โชติพัชร์ ภรณวลัย	

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

วัน / เดือน/ ปี ที่สอบ วันอังคารที่ 15 ธันวาคม 2558 เวลา 09.30 น. เป็นต้นไป

สถานที่สอบ ณ ห้อง 333 ชั้น 3 คณะเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศรับรองแล้ว

(รองศาสตราจารย์ ดร.จันทรัปรณ์ สถิตวิริยวงศ์)

คณบดีคณะเทคโนโลยีสารสนเทศ

วันที่ 21 **เดือน** ธันวาคม **พ.ศ.** 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	สแนปบอร์ด: กระดานแสดงและแบ่งปันสรุปแบบภาพ สำหรับการค้นหาเว็บไซต์แบบร่วมมือ
นักศึกษา	นาย หาญนรินทร์ เครือจิรายุส์
รหัสนักศึกษา	57606150
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
พ.ศ.	2558
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. ธีรพงศ์ ลีลานุภาพ

บทคัดย่อ

มีงานหลายประเภทที่คนอาจจำเป็นต้องค้นหาข้อมูลร่วมมือกัน เช่น การวางแผนการท่องเที่ยว การทำรายงานกลุ่ม เป็นต้น อย่างไรก็ตามระบบค้นหาในปัจจุบันยังคงถูกออกแบบเพื่อการค้นหาเพียงคนเดียว

วิทยานิพนธ์เล่มนี้จึงนำเสนอระบบค้นหาข้อมูลแบบร่วมมือรูปใหม่ ชื่อ "โคสเปช" เป็นโปรแกรมประยุกต์บนเว็บไซต์ที่ช่วยผู้ใช้ในการค้นหาร่วมกัน ซึ่งเราได้พัฒนาเครื่องมือที่สำคัญชื่อ "สแนปบอร์ด" เป็นกระดานแสดงและแบ่งปันสรุปแบบภาพ ซึ่งเป็นการนำรูปภาพมาช่วยในการสื่อสารระหว่างค้นหารูปแบบหนึ่ง โดยการสรุปแบบภาพนี้ได้จากการสร้างจากเว็บไซต์โดยผู้ค้นหาเอง ซึ่งการสรุปโดยใช้ภาพนี้ช่วยให้ผู้ใช้สามารถสื่อสารระหว่างผู้ร่วมค้นหาได้ดีกว่าการใช้ข้อความ การทดลองถูกสร้างขึ้นเพื่อจำลองเหตุการณ์ โดยให้ผู้เข้าร่วมการทดลองทำการค้นหาข้อมูลแบบคนเดียวและร่วมกันแบบต่างเวลากัน จากนั้นการปฏิสัมพันธ์ของกลุ่มผู้เข้าร่วมการทดลองระหว่างค้นหาจะถูกนำมาวิเคราะห์ เพื่อศึกษาประโยชน์ของสรุปแบบภาพในงานแต่ละประเภทซึ่งถูกยกตัวอย่างมาจากงานที่คนมักค้นหาร่วมกัน

ผลการทดลองพบว่าสรุปแบบภาพที่นำเสนอในวิทยานิพนธ์นี้ ช่วยผู้ค้นหาทั้ง 2 สถานการณ์ที่จำลอง ได้แก่ การค้นหาข้อมูลแบบคนเดียวและร่วมกันแบบต่างเวลากัน นอกจากนี้เรายังปูทางเพื่องานวิจัยในการใช้สรุปแบบภาพในอนาคต โดยทำการทดลองการสกัดคำจากสรุปแบบภาพ ซึ่งอาจจะเป็นประโยชน์ในการนำไปใช้สร้างคำแนะนำสำหรับการค้นหาแบบร่วมมือกัน เราสกัดคำจากรูปภาพโดยวิธีการ OCR (optical character recognition) และประเมินผลโดยใช้ตัวชี้วัด Precision, Recall, และ F-measure ผลการทดลองพบว่าวิธีการสกัดคำของเราให้ค่า Precision, Recall, และ F-measure สูง และสามารถนำไปใช้ประโยชน์ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis	Snapboard: Shared Space of Visual Snippets for Collaborative Web Search
Student	Mr. Hannarin Kruajirayu
Student ID.	57606150
Degree	Master of Science
Program	Information Technology
Major	Information Science
Year	2015
Thesis Advisor	Asst. Prof. Dr. Teerapong Leelanupab

ABSTRACT

Many search tasks require people to collaboratively search for information, where they often have joint information needs. Examples of these tasks are travel planning, group assignment report, and general shopping. However, such tasks are not supported by current search systems that are designed for an individual search only.

This thesis introduces a new collaborative search system, CoZpace. CoZpace is a web-based application that enables a group of users to collaborate on searching the web. We also present the main feature of CoZpace, named Snapboard, which is a shared board for a collection of group-created visual snippets. The visual snippet is a snapshot of focused and salient information captured by a user. It acts as a visual summarization of web pages, which allows any user to quickly recognize information and to revisit web pages. This thesis describes example usage scenarios and investigates the ways Snapboard facilitates users in two simulated work task situations: individual and asynchronous collaborative searches. We then analyze users' interactions and discuss how Snapboard supports search collaboration among study participants.

The evaluation showed that Snapboard helped the participants in the both simulated work task situations. Furthermore, we paved the path for future research in exploiting visual snippets to support collaborative search activities, such as generating query suggestion. We conducted a preliminary experiment to extract terms from the visual snippets by using Optical Character Recognition (OCR). Outcomes of term extraction were evaluated by using Precision, Recall, and F-measure. The result demonstrated that our method achieved high Precision, Recall and F-measure.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จลุล่วงไปได้ดี ด้วยความช่วยเหลือ ให้คำปรึกษา และสละเวลาอันมีค่าของ ผู้ช่วยศาสตราจารย์ ดร. ธีรพงศ์ ลีลานุภาพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ในการให้คำแนะนำเกี่ยวกับแนวคิดในการทำวิทยานิพนธ์ ตรวจสอบและแก้ไขข้อผิดพลาด รวมทั้งการสนับสนุนในการตีพิมพ์งานวิจัยให้ออกสู่สารธารณะชน ขอบคุณ นายนนท์ คนึงสุขเกษม สำหรับข้อแนะนำและความช่วยเหลือในการทำงานวิจัย ขอขอบคุณ Dr. Guido Zuccon ที่แนะนำและให้คำปรึกษางานวิจัย รวมทั้งช่วยเหลือและดูแลอย่างดีในการนำเสนอผลงานที่ประเทศออสเตรเลีย และขอขอบคุณ นายภาณุพงศ์ เหลืองปทุม นายเอก ตั้งสมบูรณ์ และนายศรัญ อ่อนเอื้อน ที่ให้คำแนะนำในการพัฒนาระบบ

ขอกราบขอบพระคุณคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ได้สนับสนุนทุนการศึกษาเล่าเรียนในบัณฑิตศึกษาและทุนสนับสนุนงานวิจัยทำให้การทำงานวิจัยเป็นไปอย่างราบรื่น

สุดท้ายนี้ขอกราบขอบพระคุณบิดามารดาและพี่น้องที่คอยให้การสนับสนุนด้านการศึกษาและคอยให้กำลังใจจนการเรียนผ่านพ้นไปได้ด้วยดี ขอกราบขอบพระคุณครูบาอาจารย์ทุกท่านที่ได้มอบความรู้ คุณธรรม และการใช้ชีวิตให้แก่ผู้เขียนตั้งแต่ระดับปริญญาตรีและปริญญาโทจนกระทั่งมีวันนี้ได้

หาญนรินทร์ เครือจิรายุส์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อ.....I

บทคัดย่อภาษาอังกฤษ..... II

กิตติกรรมประกาศ.....III

สารบัญ.....IV

สารบัญตาราง.....VII

สารบัญรูปภาพ.....VIII

บทที่ 1 บทนำ.....9

1.1 ที่มาและความสำคัญของวิทยานิพนธ์.....9

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....10

1.3 แนวคิดที่ใช้ในการทำวิจัย.....10

1.4 คำถามงานวิจัย.....10

1.5 การมีส่วนร่วมต่องานวิจัยในสาขาวิชา.....11

1.6 ขอบเขตของการศึกษา.....11

1.7 ขั้นตอนของการศึกษา.....11

1.8 ผลงานที่ได้รับการตีพิมพ์.....12

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....13

2.1 ทฤษฎีที่เกี่ยวข้องกับการค้นคืนข้อมูล.....13

2.2 งานวิจัยที่เกี่ยวข้องกับการค้นคืนข้อมูล.....17

บทที่ 3 โคสเปช: ระบบค้นหาข้อมูลแบบร่วมมือ.....32

3.1 สถาปัตยกรรมของระบบโคสเปช.....32

3.2 โคสเปช (CoZpace) 34

3.3 กระดานแสดงและแบ่งปันรูปภาพภายในกลุ่ม (Snapboard).....36

บทที่ 4 การศึกษาระบบโคสเปชจากกลุ่มผู้ใช้งานในเรื่องการค้นหาแบบคนเดียวและค้นหาร่วมกันแบบต่างเวลา.....38

4.1 ปัญหางานวิจัย.....38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2 ขอบเขตการศึกษา....38
4.3 สมมติฐานงานวิจัย....39
4.4 ผู้เข้าร่วมการทดลอง....39
4.5 โจทย์ในการทำการทดลอง....40
4.6 สถานการณ์จำลองในการทดลอง....40
4.7 การเก็บข้อมูลสำหรับการศึกษา....41
4.8 ขั้นตอนการทดลอง....42
4.9 ผลการทดลอง....43
บทที่ 5 การสกัดคำจากรูปภาพของหน้าเว็บเพจ....52
5.1 ปัญหางานวิจัย....52
5.2 ขอบเขตของการศึกษา....52
5.3 สมมติฐานงานวิจัย....52
5.4 ข้อตกลงเบื้องต้น....53
5.5 ข้อมูลสำหรับการทดลอง....53
5.6 Lemur search toolkit....53
5.7 การดึงคำจากรูปภาพ....54
5.8 ขั้นตอนการทดลอง....56
5.9 ผลการทดลอง....58
บทที่ 6 สรุปผลงานวิจัยและข้อเสนอแนะ....59
6.1 สรุปผลงานวิจัย....59
6.2 ข้อเสนอแนะ....60
เอกสารอ้างอิง....62
ภาคผนวก ก. ผลงานวิจัยที่ได้รับการตีพิมพ์....65
ภาคผนวก ข. เอกสารในการทำการศึกษาระบบค้นหาข้อมูลแบบร่วมมือจากกลุ่มผู้ใช้งาน....97
ประวัติผู้เขียน....109

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 กิจกรรมที่ผู้ทดลองกระทำระหว่างค้นหาข้อมูล.....22

2.2 ข้อมูลเชิงปริมาณเปรียบเทียบระบบ Cofox กับระบบ Solo และ Tog.....26

4.1 แสดงข้อมูลเชิงปริมาณที่เก็บบันทึกเพื่อการวิเคราะห์ผลการทดลอง.....42

4.2 แสดงตัวอย่างการหมุนเวียนลำดับการใช้ระบบ S1 และ S2 กับการทำโจทย์ T1, T2, T3 และ T4 ของผู้ทดลองทั้งหมด 32 คน.....43

4.3 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่ผู้ทดลองกระทำในการค้นหาข้อมูล (เฉลี่ย 16 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันในเวลาที่ต่างกัน (AC).....44

4.4 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่นักเรียนกระทำในการค้นหาข้อมูล (เฉลี่ย 16 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันในเวลาที่ต่างกัน (AC).....46

4.5 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่นักศึกษากระทำในการค้นหาข้อมูล (เฉลี่ย 16 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันในเวลาที่ต่างกัน (AC).....47

4.6 แสดงค่าเฉลี่ยคะแนนความพึงพอใจหลังทำโจทย์แต่ละข้อของผู้ทดลอง.....49

4.7 คะแนนเฉลี่ยความพึงพอใจที่มีต่อส่วนติดต่อผู้ใช้ของระบบโคสเปซหลังเสร็จสิ้นการทำการทดลองของผู้ทดลอง.....50

4.8 การเปรียบเทียบมุมมองการรับรู้ของผู้ใช้ที่มีต่อระบบหลังเสร็จสิ้นการทดลอง.....51

5.1 แสดงข้อมูลหลังทำดัชนีโดย Lemur search toolkit เพื่อใช้ในการทดลอง.....53

5.2 แสดงผลการสกัดคำจากรูปภาพของ 1,000 เว็บไซต์.....58

5.3 แสดงผลการสกัดคำจากรูปภาพแบ่งตามกลุ่มของจำนวนคำที่มีในเว็บไซต์.....58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูปที่ หน้า

2.1 ตัวอย่างระบบค้นหาข้อมูล CoSearch บนคอมพิวเตอร์ตั้งโต๊ะ....18

2.2 ตัวอย่างระบบค้นหาข้อมูล CoSearch บนโทรศัพท์มือถือ....18

2.3 ตัวอย่างระบบค้นหาข้อมูล ViGOR....20

2.4 ตัวอย่างระบบค้นหาข้อมูล CollabSearch....21

2.5 เปอร์เซ็นต์ของคู่การกระทำใน 3 เงื่อนไข....22

2.6 แสดงปริมาณกิจกรรมที่ผู้ทดลองทำก่อนการใส่คำค้นหาข้อมูลของทั้ง 3 รูปแบบการค้นหา....23

2.7 ตัวอย่างระบบค้นหาข้อมูล SearchTogether....24

2.8 ตัวอย่างระบบค้นหาข้อมูล WeSearch....25

2.9 ตัวอย่างระบบค้นหาข้อมูล Cofox....26

2.10 ตัวอย่าง (ล่าง) เว็บไซต์ และ (บน) สรุปแบบภาพ....27

2.11 ตัวอย่าง (a) รูปภาพที่โดดเด่นที่ดึงได้จากเว็บไซต์ (b) สรุปแบบภาพและ(c) รูปภาพ ขนาดย่อของเว็บไซต์จริง....28

2.12 ตัวอย่างการนำรูปภาพขนาดย่อมารวมกับสรุปข้อความตัวอักษรของเว็บไซต์....29

2.13 ตัวอย่างการแนะนำรูปภาพเพื่อการค้นหาข้อมูล....30

2.14 ตัวอย่างการแนะนำวีดีโอเพื่อการค้นหาข้อมูล....30

2.15 ตัวอย่างการแนะนำรูปภาพและองค์ประกอบภายในรูปภาพเพื่อการค้นหาข้อมูล....31

3.1 ตัวอย่างสถาปัตยกรรมของระบบโคสเปช....32

3.2 หน้าตาระบบค้นหาข้อมูล CoZpace....34

3.3 หน้าต่างแสดงผลหลังกดปุ่ม Relevance....35

3.4 หน้าต่างแสดงผลหลังกดปุ่ม Queries....35

3.5 หน้าต่างแสดงผลหลังกดปุ่ม Views....35

3.6 ตัวอย่าง (1) การตัดสรุปแบบภาพโดยผู้ใช้ (2) การทำสรุปแบบภาพโดยการรวมกัน ของภาพที่ผู้ใช้ตัดและข้อความสรุปเว็บไซต์และตัวอักษร....37

3.7 ตัวอย่างฟังก์ชัน สแนปบอร์ด กระดานแสดงและแบ่งปันรูปภาพภายในกลุ่ม....37

4.1 ผู้เข้าร่วมการทดลองเป็นนักเรียนจากโรงเรียนนครนายกวิทยาคม....39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ (ต่อ)

รูปที่ หน้า

5.1 วิธีการสร้างคำแนะนำสำหรับการค้นหา....54
5.2 ภาพจำลองการทำงานของ Wayback tool....55
5.3 การขยายภาพให้ใหญ่ขึ้นจากเดิมสามเท่าและปรับภาพให้เป็นภาพขาว-ดำ....55
5.4 ภาพจำลองขั้นตอนในการทำทดลอง....56
5.5 ภาพจำลองเซตของคำ T1 T2 และ T3 สำหรับการทดลอง....57

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของวิทยานิพนธ์

เทคโนโลยีที่ใช้ในการค้นหาข้อมูลได้ถูกพัฒนาและมีความก้าวหน้าขึ้นอย่างมาก แต่ระบบค้นหาข้อมูล (search engine) ที่นิยม เช่น Google หรือ Bing ยังคงถูกออกแบบมาเพื่อรองรับการค้นหาข้อมูลเพียงคนเดียว (individual search) แม้ว่าจะมีการศึกษาและพบว่ามีงานหลายประเภทที่คนเรามักค้นหาร่วมกันบ่อย ๆ ก็ตาม [1] ตัวอย่างเช่น การวางแผนการท่องเที่ยว การซื้อของที่มีราคาแพงที่ต้องการผู้เชี่ยวชาญในการให้คำปรึกษา และการทำรายงานกลุ่ม เป็นต้น พวกเขาจำเป็นต้องปรึกษาและค้นหาข้อมูลร่วมกัน โดยใช้ระบบค้นหาข้อมูลร่วมกับระบบที่ช่วยในการสื่อสาร เช่น เครือข่ายสังคม (social network) แม้จะมีระบบค้นหาข้อมูลแบบร่วมมือ (collaborative search system) ถูกพัฒนาขึ้น แต่ก็ยังไม่เป็นที่นิยมเท่าที่ควร โดยระบบค้นหาข้อมูลแบบร่วมมือในปัจจุบัน จะช่วยผู้ค้นหา (searcher) ในการสื่อสารร่วมกันผ่านการแบ่งปัน (share) ข้อมูลในรูปแบบข้อความตัวอักษรเป็นส่วนใหญ่ (text-based communication) เช่น การบันทึกเว็บไซต์ที่เกี่ยวข้อง (bookmarks) การสนทนา (instant messaging) หรือการแสดงความคิดเห็นเว็บไซต์ (comment) เป็นต้น แม้จะมีการนำภาพมาช่วยในติดต่อสื่อสาร (visual-based communication) แต่ก็ยังต้องค้นหาข้อมูลในเวลาเดียวกัน (synchronous collaboration) เท่านั้น เช่น การแสดงวิดีโอภาพการค้นหาของผู้ร่วมค้นหา (collaborator) แบบเรียลไทม์

วิทยานิพนธ์เล่มนี้นำเสนอการนำสรุปแบบภาพ (visual snippet) มาช่วยในการสื่อสารแบบต่างเวลากัน (asynchronous collaboration) ระหว่างการค้นหาข้อมูลแบบร่วมมือ โดยเราได้ทำการปรับปรุงและพัฒนาระบบค้นหาข้อมูลแบบร่วมมือกันขึ้นชื่อ "โคสเปซ" (CoZpace) [2] [3] และคิดค้นฟังก์ชันใหม่ชื่อ "สแนปบอร์ด กระดานแสดงและแบ่งปันสรุปแบบภาพภายในกลุ่ม" (Snapboard) ซึ่งเป็นการนำรูปภาพที่ตัดจากเว็บไซต์โดยผู้ใช้เอง มาช่วยในการติดต่อสื่อสารระหว่างค้นหาข้อมูล โดยวิทยานิพนธ์เล่มนี้มุ่งเน้นที่จะแนะนำระบบค้นหาข้อมูลแบบร่วมมือที่มีประสิทธิภาพมากขึ้น โดยทำการศึกษาโคสเปซ ระบบค้นหาข้อมูลแบบร่วมมือจากกลุ่มผู้ใช้งานจริง [4] ซึ่งเราจะเก็บข้อมูลทั้งในเชิงปริมาณและเชิงคุณภาพ โดยข้อมูลในเชิงปริมาณ ระบบจะมีการเก็บข้อมูล (log) ของกิจกรรมระหว่างการค้นหาของผู้ทำการทดลอง เพื่อนำมาวิเคราะห์ประสิทธิภาพของฟังก์ชันสแนปบอร์ด ซึ่งเป็นฟังก์ชันใหม่ที่เราคิดค้นขึ้น และเปรียบเทียบฟังก์ชันสแนปบอร์ดกับฟังก์ชันแสดงความคิดเห็นที่มีอยู่ในระบบปัจจุบัน ในส่วนของข้อมูลเชิงคุณภาพเราจะใช้แบบสอบถามแบบให้คะแนน (5 point Likert-scale

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

questionnaire) ถามผู้ทดลองในเรื่องของประสิทธิภาพของแต่ละฟังก์ชันและให้คะแนน รวมทั้งมีคำถามปลายเปิดเพื่อให้พวกเขาเสนอความคิดเห็น นอกจากนี้เรายังนำเสนอการใช้ประโยชน์จากสรุปแบบภาพ โดยการสกัดคำจากรูปภาพสรุปของหน้าเว็บไซต์ด้วยวิธีการ OCR (optical character recognition) [5] เพื่อเป็นแนวทางในสร้างคำแนะนำสำหรับการค้นหา (query suggestion) ภายในกลุ่ม

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. นำเสนอฟังก์ชันใหม่ได้แก่ สแนปบอร์ด ซึ่งช่วยในการค้นหาข้อมูลแบบร่วมมือ
2. เพิ่มการรับรู้ของผู้ค้นหา โดยการนำสรุปแบบภาพมาช่วยในการสื่อสารแทนข้อความตัวอีกษร
3. ประเมินประสิทธิภาพของ สแนปบอร์ด ในการค้นหาข้อมูลแบบต่างเวลากัน
4. เป็นแนวทางในการนำสรุปแบบภาพมาใช้งานในด้านการค้นหาข้อมูลแบบร่วมมือและการสร้างคำแนะนำสำหรับการค้นหาเบื้องต้น

1.3 แนวคิดที่ใช้ในการทำวิจัย

เราทำการปรับปรุงและพัฒนาระบบค้นหาข้อมูลแบบร่วมมือกันขึ้นชื่อ “โคสเปช” และคิดค้นฟังก์ชันใหม่ชื่อ “กระดานแสดงและแบ่งปันรูปภาพเฉพาะกลุ่ม” ซึ่งเป็นการนำรูปภาพมาช่วยในการสื่อสาร โดยเป็นการนำรูปภาพสรุปของหน้าเว็บไซต์ ที่ผู้ค้นหาภายในกลุ่มทำการตัด (snapshot) จากหน้าเว็บไซต์เอง ซึ่งน่าจะตรงกับความต้องการภายในกลุ่ม มาแบ่งปันให้ผู้ร่วมค้นหา เพื่อเป็นการปรับปรุงการรับรู้และการสื่อสารระหว่างการค้นหาข้อมูลร่วมกันภายในกลุ่มให้ดียิ่งขึ้น ซึ่งไม่เพียงแค่สนับสนุนการค้นหาข้อมูลร่วมกันในเวลาเดียวกันเท่านั้น แต่ยังสนับสนุนการค้นหาข้อมูลร่วมกันแบบต่างเวลากันอีกด้วย นอกจากนี้ในรูปภาพสรุปบางส่วนของเว็บไซต์ที่ผู้ค้นหาตัดมานี้ ยังอาจจะมีคำสำคัญ (keyword) ซึ่งน่าจะนำมาใช้ประโยชน์ในการสร้างคำแนะนำสำหรับค้นหา (query suggestion) ได้ โดยการดึงคำจากรูปภาพสรุปของหน้าเว็บไซต์มาทำการประมวลผลภาษาธรรมชาติ (natural language processing) ในการสร้างคำแนะนำสำหรับค้นหา

1.4 คำถามงานวิจัย

1. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ค้นหาในการค้นหางานที่คนมักค้นหาร่วมกัน
2. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ใช้ในการค้นหาคนเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ใช้ในการค้นหาร่วมกันแบบต่างเวลา
4. การสกัดคำจากสรุปแบบภาพ ให้ความถูกต้องที่น่าเชื่อถือได้
5. จำนวนคำในสรุปแบบภาพ ไม่มีผลต่อวิธีการสกัดคำ

1.5 การมีส่วนร่วมต่องานวิจัยในสาขาวิชา (Contribution)

1. งานวิจัยนี้ได้นำเสนอฟังก์ชันใหม่ชื่อ "สแนปบอร์ด กระดานแสดงและแบ่งปันสรุปแบบภาพภายในกลุ่ม" ซึ่งเป็นการนำรูปภาพมาช่วยในการสื่อสารแทนข้อความตัวอักษร เพื่อปรับปรุงระบบค้นหาข้อมูลแบบร่วมมือ
2. ทำการทดลองกับกลุ่มผู้ใช้ เพื่อศึกษาว่าฟังก์ชันสแนปบอร์ดช่วยผู้ใช้ในการค้นหาข้อมูลแบบร่วมมือกันหรือไม่
3. นำเสนอการใช้ประโยชน์จากสรุปแบบภาพ โดยการสกัดคำจากสรุปแบบภาพเบื้องต้นเพื่อนำไปใช้สร้างคำแนะนำสำหรับการค้นหา
4. ทำการทดลองเพื่อตรวจสอบความน่าเชื่อถือของวิธีการในการสกัดคำจากสรุปแบบภาพ โดยการเปรียบเทียบกับเครื่องมือในปัจจุบัน

1.6 ขอบเขตของการศึกษา

1. ผลการค้นหาเป็นข้อมูลประเภทเว็บไซต์เท่านั้น
2. ศึกษาในเรื่องส่วนต่อประสานกับผู้ใช้ (user interface)
3. ศึกษาระบบโคสเปช ในเรื่องการร่วมมือกันในการค้นหาข้อมูลแบบตั้งใจของผู้ใช้ (explicit collaboration) และต่างสถานที่ (distribute)
4. ใช้ผลการค้นหาของ Microsoft Bing ในการศึกษาฟังก์ชันสแนปบอร์ดจากกลุ่มผู้ใช้
5. ศึกษากลุ่มผู้ใช้ทั้งในการค้นหาข้อมูลแบบคนเดียว (individual search) และแบบร่วมมือกันในเวลาที่ต่างกัน
6. เก็บข้อมูลทั้งในรูปเชิงปริมาณและเชิงคุณภาพเพื่อวิเคราะห์ผลการทดลอง
7. ใช้ข้อมูลของ Clueweb09 ในการทำการทดลองการสกัดคำจากสรุปแบบภาพ

1.7 ขั้นตอนของการศึกษา

1. ศึกษาระบบค้นหาข้อมูลร่วมกันในปัจจุบันและทฤษฎีที่เกี่ยวข้องกับระบบ
2. ศึกษาการปรับปรุงการสรุปเว็บไซต์ที่พัฒนาจากรูปแบบข้อความตัวอักษรมาเป็นรูปภาพ เพื่อประยุกต์มาใช้ในการสื่อสารในระบบค้นหาข้อมูลแบบร่วมมือ
3. ออกแบบขั้นตอนและวิธีการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. เลือกกลุ่มเป้าหมายที่ใช้ในการศึกษา
5. ทำการทดลองกับกลุ่มเป้าหมาย
6. วิเคราะห์ผลการทดลอง
7. อภิปรายและเปรียบเทียบผลการวัดประสิทธิภาพ
8. สรุปผลการวิจัย

1.8 ผลงานที่ได้รับการตีพิมพ์

ผลงานตีพิมพ์ตลอดการทำวิทยานิพนธ์นี้มี 4 บทความ (สามารถดูได้ที่ภาคผนวก ก)

1. การพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม
   Hannarin Kruajirayu, Ake Tangsomboon, and Teerapong Leelanupab; in the National 6th Conference on Information Technology, NCIT 2014; Thailand.
2. CoZpace: A Proposal for Collaborative Web Search for Sharing Search Records and Interactions.
   Hannarin Kruajirayu, Ake Tangsomboon, and Teerapong Leelanupab; in the 3rd ICT International Student Project Conference, ICT-ISPC 2014; Thailand.
3. Extracting Visual Snippets for Query Suggestions in Collaborative Web Search.
   Hannarin Kruajirayu, and Teerapong Leelanupab; in Proceedings of the 20th International Symposium on Artificial Life and Robotics, AROB 2015; Beppu, Japan.
4. Snapboard: A Shared Space of Visual Snippets. A Study in Individual and Asynchronous Collaborative Web Search.
   Teerapong Leelanupab, **Hannarin Kruajirayu**, and Nont Kanungsukkasem; in Proceedings of the Asia Information Retrieval Societies Conference, AIRS 2015; Brisbane, Queensland, Australia.

ส่วนที่เหลือของวิทยานิพนธ์เล่มนี้ได้แก่ บทที่ 2 ซึ่งจะอธิบายเกี่ยวกับทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการค้นคืน บทที่ 3 นำเสนอระบบโคสเปช ซึ่งเป็นระบบค้นหาข้อมูลแบบร่วมมือที่คิดขึ้น บทที่ 4 การศึกษาระบบโคสเปชจากกลุ่มผู้ใช้งานในเรื่องการค้นหาแบบคนเดียวและค้นหาร่วมกันแบบต่างเวลา บทที่ 5 การสกัดคำจากรูปภาพของหน้าเว็บเพจ บทที่ 6 สรุปผลงานวิจัยและข้อเสนอแนะ และบทที่7 ภาคผนวก (ผลงานที่ตีพิมพ์และแบบสอบถาม)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทที่ 2 จะอธิบายเกี่ยวกับทฤษฎีและงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์เล่มนี้ โดยมีดังนี้ คือ 1) ทฤษฎีที่เกี่ยวข้องกับการค้นคืนข้อมูล ได้แก่ การค้นคืนแบบโต้ตอบ การประเมินระบบค้นหาข้อมูล และการจำแนกมิติของการร่วมมือกัน และ 2) งานวิจัยที่เกี่ยวข้องกับการค้นคืนข้อมูล ได้แก่ ระบบค้นหาข้อมูลแบบร่วมมือในปัจจุบัน การสรุปแบบภาพ และคำแนะนำสำหรับการค้นหา

2.1 ทฤษฎีที่เกี่ยวข้องกับการค้นคืนข้อมูล

2.1.1 ทฤษฎีการค้นคืนข้อมูลแบบโต้ตอบ (Interactive Information Retrieval)

การค้นคืนข้อมูลแบบโต้ตอบเป็นส่วนหนึ่งในงานด้าน Human-Computer Information Retrieval (HCIR) คือการศึกษาเทคนิคการดึงข้อมูลที่ทีมีการนำมนุษย์เข้าร่วมการทดลองในกระบวนการค้นหา เป็นการมุ่งเน้นไปที่การปรับปรุงความสามรถของระบบในการนำเสนอผลการค้นหาบนส่วนติดต่อผู้ใช้ (user interface) และมีการเปลี่ยนแปลงผลการค้นหาที่เกี่ยวข้องกับการความต้องการที่เปลี่ยนแปลงไปของกลุ่มผู้ใช้ [6] ดังนั้นส่วนสำคัญในงานวิจัยด้านการค้นคืนข้อมูลแบบโต้ตอบคือ การสร้างส่วนติดต่อผู้ใช้แบบโต้ตอบ (interactive user interface) และการมีส่วนร่วมของมนุษย์ในกระบวนการค้นหา

2.1.2 ทฤษฎีการประเมินระบบค้นหาข้อมูล (Evaluation in Information Retrieval)

เป้าหมายของการค้นคืนข้อมูล (Information retrieval) คือ ระบบค้นหาควรที่จะคืนข้อมูลให้ตรงกับที่ผู้ค้นหาต้องการ (information need) การประเมินระบบค้นคืนจึงเป็นสิ่งสำคัญส่วนหนึ่งในงานวิจัยด้านการค้นคืนข้อมูล เป้าหมายหมายหลักคือการวัดประสิทธิภาพและประสิทธิผลของระบบค้นหา ตัวอย่างของการวัดประสิทธิภาพ เช่น ระบบสามารถคืนผลการค้นหาได้เร็วหรือไม่ และตัวอย่างของการวัดประสิทธิผล เช่น การพิจารณาว่าผลการค้นหาตอบสนองต่อความต้องการของผู้ใช้หรือไม่ ในการทดลองด้านการค้นคืนข้อมูลส่วนใหญ่มีเป้าหมายในการประเมินประสิทธิผลของระบบค้นหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสิทธิผลของระบบค้นหาสามรถสรุปได้จากความพึงพอใจของผู้ใช้ที่มีต่อระบบค้นหา การประเมินระบบค้นหาข้อมูลสามารถแบ่งออกได้เป็น 2 ประเภทคือ 1) การประเมินผลโดยมุ่งเน้นระบบค้นหาเป็นศูนย์กลาง 2) การประเมินผลโดยมุ่งเน้นกลุ่มผู้ใช้เป็นศูนย์กลาง ซึ่งจะอธิบายดังต่อไปนี้

1. การประเมินผลโดยมุ่งเน้นระบบค้นหาเป็นศูนย์กลาง (System-Oriented Evaluation)

วิธีการประเมินการค้นคืนข้อมูลส่วนใหญ่คือ การประเมินผลโดยมุ่งเน้นระบบค้นหาเป็นศูนย์กลาง จุดเริ่มต้นของการประเมินผลระบบค้นคืนแบบนี้มาจากงานของ Cleverdon และคณะ (1966) ผู้ที่นำเสนอชุดทดสอบ (test collection) ซึ่งถูกใช้ในการตั้งค่าและควบคุมการประเมินผลระบบค้นคืนโดยใช้คอมพิวเตอร์ วิธีการนี้โดยทั่วไปจะเรียกว่าเป็นการประเมินผลแบบกระบวนการแคลนฟิล (Cranfield evaluation paradigm) [7] ในกระบวนการแคลนฟิลนี้ นักวิจัยจะต้องรวบรวมชุดทดสอบหรือคลังประเมิน (evaluation corpus) [8] ซึ่งประกอบด้วย เอกสาร คำที่ใช้ค้นหา และการตัดสินความเกี่ยวข้องของเอกสาร รวมถึงตัวชี้วัด Precision และ Recall โดยแนวคิดแบบแคนฟิลมีดังนี้คือ

- การรวบรวมเอกสาร
- สร้างชุดข้อมูลของความต้องการ คือการเก็บคำที่ใช้ค้นหา ซึ่งบางครั้งอาจมาพร้อมกับคำอธิบายสั้น ๆ ที่สัมพันธ์กับคำที่ใช้ค้นหานั้น
- การรวบรวมการตัดสินความเกี่ยวข้องของความต้องการของผู้ใช้กับเอกสารที่ระบบค้นหาคืนให้ โดยปกติจะตัดสินเอกสารเป็น 2 ประเภทคือ 1) เอกสารที่เกี่ยวข้อง (Relevant document) และ 2) เอกสารที่ไม่เกี่ยวข้อง (non-relevant document) หรือตัดสินความเกี่ยวข้องของเอกสารแบบมีหลายระดับ เช่น ให้เลือกว่าเอกสารมีความเกี่ยวข้องในระดับไหน โดยให้เลือกตั้งแต่ 0-4 (0 คือเอกสารไม่เกี่ยวข้อง และ 4 คือเอกสารเกี่ยวข้องมากสุด) ซึ่งความเกี่ยวข้องของเอกสารเหล่านี้ จะนำไปใช้เพื่อประเมินผลการค้นหาที่คืนมาจากระบบค้นหา
- การประเมินระบบค้นหาโดยใช้การตัดสินความเกี่ยวข้องจากตัวชี้วัดต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การประเมินผลโดยมุ่งเน้นกลุ่มผู้ใช้เป็นศูนย์กลาง (User-Centered Evaluation)

ความต้องการสามารถเปลี่ยนแปลงได้ในระหว่างค้นหาข้อมูล Ingwersen และ Jarvelin (2005) พบความเป็นจริงว่า ความต้องการของผู้ค้นหามีการเปลี่ยนแปลงอยู่ตลอดเวลา เมื่อผู้ค้นหาได้รับข้อมูลใหม่จากผลการค้นหา [9] Robertson และ Hancock-Beaulieu (1992) ให้เหตุผลว่า "การประเมินผลโดยมุ่งเน้นระบบค้นหาเป็นศูนย์กลาง" ไม่เหมาะกับการประเมินผลระบบค้นคืนที่มีการโต้ตอบกับผู้ค้นหา (interactive IR system) [10] Borlund (2003) แนะนำว่าควรนำผู้ใช้มาเป็นอาสาสมัครในการทำทดลองเพื่อประเมินผลระบบค้นคืน [11] วิธีการของเธอมองการรับรู้และพฤติกรรมของผู้ใช้เป็นศูนย์กลางในการประเมินผลแทนการมองประสิทธิผลของระบบค้นคืนโดยใช้ตัวชี้วัดทั่วไปเช่น Precision และ Recall วิธีการโดยทั่วไปที่ใช้ในการรวบรวมข้อมูลในการประเมินผลโดยมุ่งเน้นผู้ใช้เป็นศูนย์กลางมีดังนี้คือ

- การเก็บ logs ข้อมูลกิจกรรมที่ผู้ค้นหากระทำระหว่างค้นหาข้อมูล
- การทำแบบสอบถามก่อน-หลังทำการทดลอง
- การสัมภาษณ์เกี่ยวกับประสบการณ์ในการค้นหาข้อมูล
- การทำวีดีโอสังเกตพฤติกรรมการค้นหา

2.1.3 ทฤษฎีการจำแนกมิติของการร่วมมือกันระหว่างการค้นหาข้อมูล (Taxonomy of Collaboration)

เราสามารถจำแนกมิติของการร่วมมือในการค้นหาข้อมูลในปัจจุบันได้เป็น 4 มิติ ดังนี้คือ [12]

1. ความตั้งใจในการร่วมมือกัน (Intent)

การร่วมมือกันโดยนัย (implicit collaboration) คือ ผู้ใช้ไม่ได้รับรู้ว่าตัวเองกำลังร่วมมือในการค้นหาข้อมูลกับคนอื่นอยู่ เพราะผลการค้นหาถูกปรับปรุงเนื่องจาก ระบบค้นหาข้อมูลมีการเก็บประวัติการค้นหา หรือพฤติกรรมของกลุ่มคน เพื่อนำไปปรับปรุงผลการค้นหา

การร่วมมือกันอย่างชัดเจน (explicit collaboration) คือ กลุ่มผู้ใช้จำนวนน้อยที่ตั้งใจที่จะค้นหาข้อมูลร่วมกัน ซึ่งระบบจะมีเครื่องมือให้ผู้ใช้สามารถร่วมมือกันค้นหาข้อมูลอย่างชัดเจน ทำให้ผู้ใช้รับรู้ว่าพวกเขาค้นหาข้อมูลร่วม ตัวอย่างเครื่องมือที่ช่วยในการร่วมมือกันค้นหาเช่น ฟังก์ชันการสนทนา แบ่งปันเว็บไซต์ที่เกี่ยวข้อง และแสดงความคิดเห็นในเว็บไซต์ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ความลึกของการเป็นตัวกลางของระบบที่ช่วยในการร่วมมือกันค้นหาข้อมูล (Depth of Mediation)

ส่วนต่อประสานกับผู้ใช้ (user interface mediation) คือ ระบบค้นหาข้อมูลมีเพียงส่วนต่อประสานผู้ใช้เท่านั้น ที่ใช้เป็นตัวกลางในการติดต่อสื่อสาร

อัลกอริทึม (algorithmic mediation) คือ ระบบค้นหาข้อมูลที่นอกจากจะมีส่วนต่อประสานผู้ใช้แล้ว ยังมีการนำอัลกอริทึมมาช่วยในประมวลผลผลการค้นหาของผู้ใช้ เพื่อนำไปปรับปรุงผลการค้นหาให้ผู้ใช้คนอื่นที่ค้นหาข้อมูลในเรื่องเดียวกัน ตัวอย่างเช่น ระบบแนะนำ (recommender system)

3. การร่วมมือโดยคำนึงถึงความพร้อมเพียงกัน (Concurrency)

การร่วมมือกันในเวลาเดียวกัน (synchronous collaboration) คือ ผู้ใช้สามารถค้นหาข้อมูลร่วมกันในเวลาเดียวกันได้ โดยระบบค้นหามีเครื่องมือช่วยให้ผู้ใช้ทำงานร่วมกัน ตัวอย่างเช่น การสนทนา (instant message) และการประชุมทางไกลด้วยวีดีโอ (video conference) เป็นต้น

การร่วมมือกันในเวลาที่ต่างกัน (asynchronous collaboration) คือ ผู้ใช้สามารถค้นหาข้อมูลร่วมกันในเวลาที่ต่างกันได้ โดยระบบจะมีการบันทึกผลการค้นหา เช่น คำที่ใช้ในการค้นหา เว็บไซต์ที่เกี่ยวข้อง และประวัติการสนทนา เป็นต้น เพื่อแสดงผลการค้นหาให้ผู้ใช้คนอื่นที่ค้นหาร่วมกันได้ดูย้อนหลัง เมื่อพวกเขาไม่ได้ค้นหาในเวลาเดียวกัน

4. สถานที่ในการร่วมมือกัน (Location)

ค้นหาข้อมูลในสถานที่เดียวกัน (co-located) คือ ผู้ใช้สามารถค้นหาข้อมูลร่วมกันในสถานที่เดียวกันและในเวลาเดียว โดยจะใช้คอมพิวเตอร์เครื่องเดียวกันหรือต่างเครื่องกันก็ได้ แต่พวกเขาสามารถค้นหาข้อมูลร่วมกันได้

ค้นหาข้อมูลในสถานที่ต่างกัน (distributed) คือ ผู้ใช้สามารถค้นหาข้อมูลร่วมกันในระยะไกลหรือต่างสถานที่กันได้นั้นเอง โดยระบบค้นหาข้อมูลจะมีเครื่องมือที่ช่วยผู้ใช้ในการติดต่อสื่อสารระหว่างทำการค้นหาข้อมูล ไม่ว่าจะค้นหาข้อมูลในเวลาเดียวกันหรือต่างเวลากัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

2.2 งานวิจัยที่เกี่ยวข้องกับการค้นคืนข้อมูล

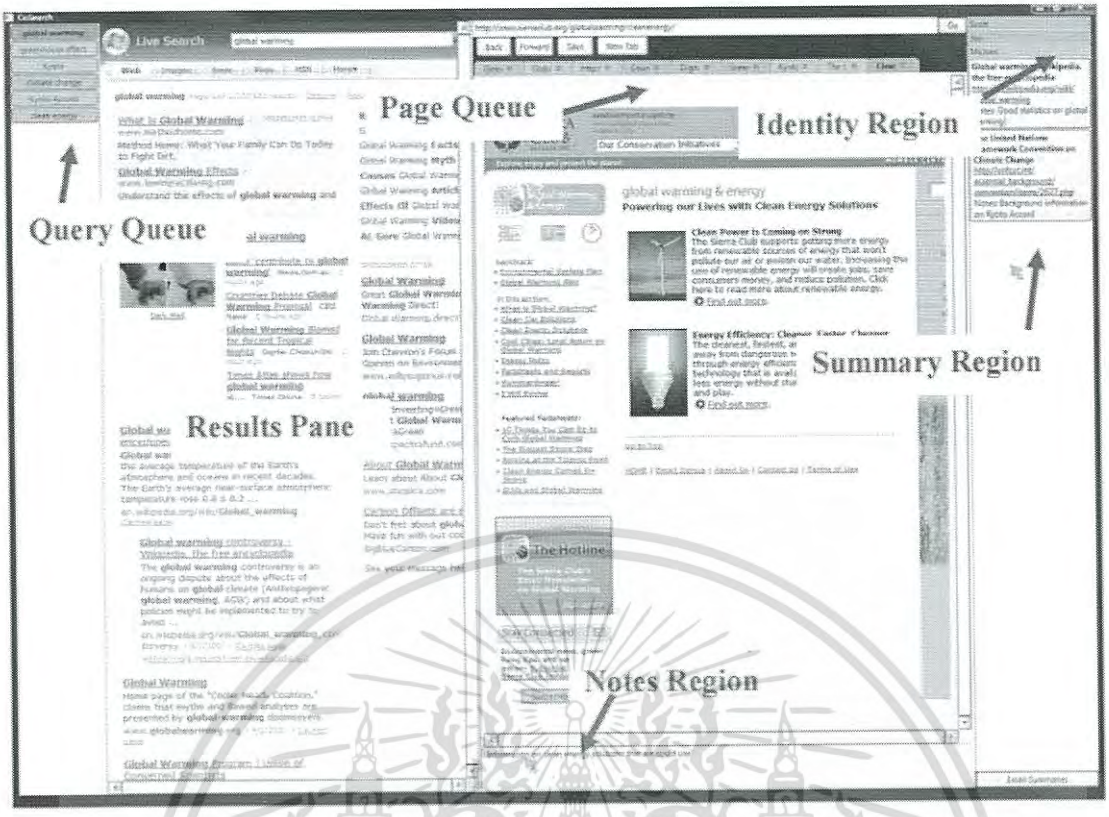
2.2.1 งานวิจัยที่เกี่ยวข้องกับระบบค้นหาข้อมูลแบบร่วมมือในปัจจุบัน (Collaborative Search System)

1. CoSearch

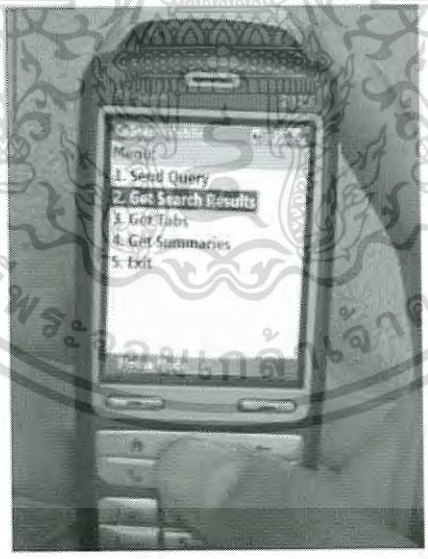
CoSearch ถูกพัฒนาโดย S. Amershi และ M. R. Morris (2008) พวกเขานำอุปกรณ์รอบ ๆ ตัวมาช่วยในการค้นหาข้อมูลแบบร่วมมือกัน ได้แก่ มือถือ และเมาส์ [13] เนื่องจากพวกเขาพบปัญหาว่า ในสถานที่ซึ่งมีทรัพยากรอย่างจำกัด เช่น ห้องสมุดในโรงเรียนที่มีเครื่องคอมพิวเตอร์น้อย หรือครอบครัวที่มีสมาชิกหลายคนแต่มีคอมพิวเตอร์เพียงเครื่องเดียว ทำให้กลุ่มผู้ค้นหามักจะยืนล้อมรอบคอมพิวเตอร์เครื่องเดียวในการทำการค้นหาข้อมูลร่วมกันบ่อย ๆ พวกเขาจึงพัฒนาระบบ CoSearch ซึ่งเป็นระบบที่สามารถค้นหาข้อมูลได้หลายคนในคอมพิวเตอร์เครื่องเดียวกัน โดยการใช้เมาส์ได้หลายตัว ซึ่งมีการระบุตัวบุคคลโดยแยกตามเมาส์ตามสี ซึ่งแสดงในรูปที่ 2.1 และพวกเขายังพัฒนาระบบค้นหาข้อมูลให้สามารถทำงานได้บนโทรศัพท์มือถือ เพื่อค้นหาข้อมูลร่วมกันกับเพื่อนที่ทำงานบนคอมพิวเตอร์ได้ เพราะพวกเขาคิดว่ามือถือเป็นอุปกรณ์ที่อยู่รอบ ๆ ตัวที่ทุกคนมีในปัจจุบัน โดยหน้าตาระบบก็จะแตกต่างจากคอมพิวเตอร์ ซึ่งถูกออกแบบมาเพื่อค้นหาข้อมูลบนโทรศัพท์มือถือโดยเฉพาะ ซึ่งแสดงในรูปที่ 2.2

ในการทดลองพวกเขาเปรียบเทียบระบบ CoSearch กับระบบค้นหาข้อมูลทั่วไป ในสถานการณ์ที่มีคอมพิวเตอร์เครื่องเดียว พวกเขาพบว่าระบบ CoSearch ช่วยผู้ใช้ในการค้นหาข้อมูลได้ดีกว่า โดยได้ข้อมูลจากการทำแบบสอบถามแบบให้คะแนน นอกจากนี้พวกเขายังเปรียบเทียบระบบ CoSearch กับระบบค้นหาข้อมูลทั่วไป ในสถานการณ์ที่มีคอมพิวเตอร์หลายเครื่อง เพื่อให้มั่นใจว่าระบบ CoSearch ไม่ได้ลดประโยชน์ในการสื่อสารระหว่างการค้นหาข้อมูลร่วมกัน เมื่อมีคอมพิวเตอร์หลายเครื่อง พวกเขาพบว่า กลุ่มผู้ใช้รู้สึกดีกว่ากับการได้ใช้คอมพิวเตอร์แยกกันในการค้นหาข้อมูลแบบร่วมมือเมื่อเปรียบเทียบกับใช้คอมพิวเตอร์เครื่องเดียวแม้จะมีระบบ CoSearch ก็ตาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
143970



รูปที่ 2.1 ตัวอย่างระบบค้นหาข้อมูล CoSearch บนคอมพิวเตอร์ตั้งโต๊ะ



รูปที่ 2.2 ตัวอย่างระบบค้นหาข้อมูล CoSearch บนโทรศัพท์มือถือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ViGOR

ViGOR ถูกพัฒนาโดย M. Halvey และคณะ (2010) ซึ่งเป็นระบบค้นหาข้อมูลแบบร่วมมือกันที่แตกต่างจากระบบทั่วไปคือ ViGOR ถูกออกแบบเพื่อการค้นหาข้อมูลประเภทวีดีโอในสถานการณ์ที่ต้องค้นหาร่วมกัน เนื่องจากข้อมูลประเภทวีดีโอเป็นที่นิยมในการค้นหามากขึ้น โดย ViGORออกแบบมารองรับการร่วมมือกันในเวลาที่ต่างกัน และร่วมมือกันแบบตั้งใจ [14] รูปที่ 2.3 แสดงตัวอย่างของระบบ ViGOR โดยระบบแบ่งเป็น 3 ส่วนดังนี้คือ

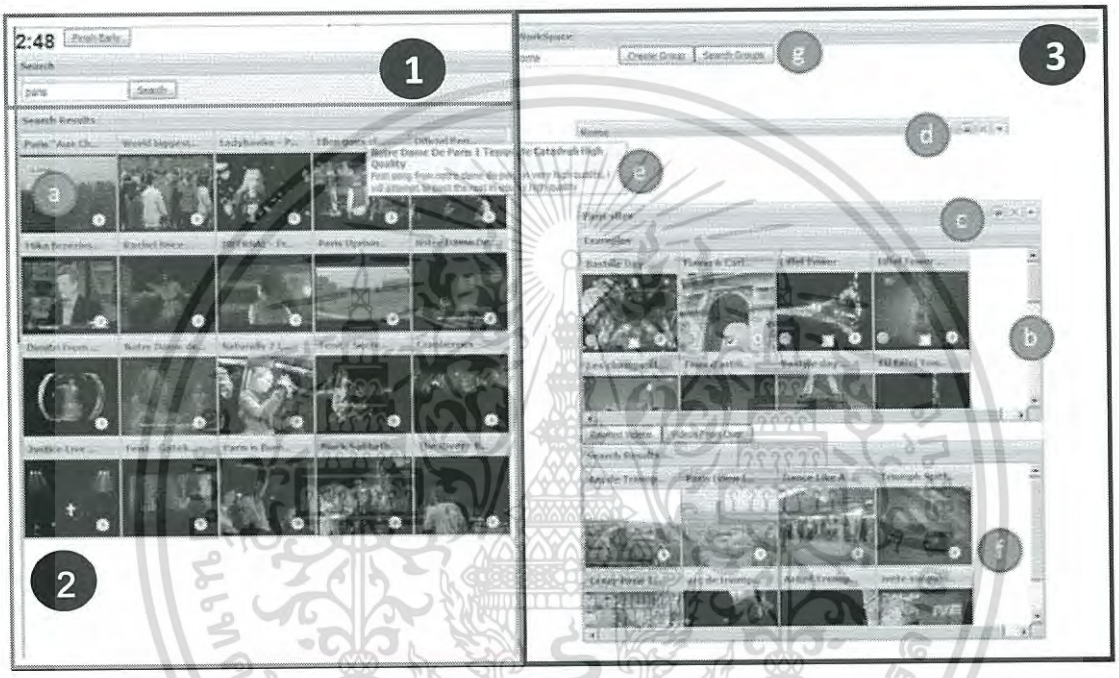
- แถบค้นหา (search panel) อยู่ด้านบนขวา สำหรับใส่ชื่อวีดีโอที่ต้องการค้นหา
- พื้นที่แสดงผลการค้นหา (result display area) โดย ViGOR ใช้ผลการค้นหาวีดีโอของ YouTube API และสามารถอ่านรายละเอียดวีดีโอได้โดยการวางเมาส์บนวีดีโอนั้น 1 วินาที
- พื้นที่สำหรับทำงานร่วมกัน (workspace) ใช้ในแบ่งปันวีดีโอร่วมกับเพื่อน ซึ่งผู้ค้นหาสามารถสร้างกลุ่มโดยการตั้งชื่อที่ต้องการค้นหาและแบ่งปันวีดีโอภายในกรุ๊ปได้ โดยการลากวีดีโอที่เกี่ยวข้องจากผลการค้นหามาทางด้านซ้ายของตัวเอง ใส่ในพื้นที่สำหรับทำงานร่วมกันทางด้านขวาในกลุ่มที่สร้างขึ้น โดยระบบจะบอกว่าวีดีโอนี้มาจากใครและสามารถให้คำอธิบายประกอบเพิ่มได้

เป้าหมายในการทดลองพวกเขาต้องการตรวจสอบว่าเครื่องมือต่าง ๆ ใน ViGOR ช่วยในการค้นหาวีดีโอแบบร่วมมือหรือไม่ โดยพวกเขาจำลองเหตุการณ์ในการค้นหาวีดีโอร่วมกันออกเป็น 3 แบบ ดังนี้คือ

- Start Scenario คือผู้ทดลองจะเริ่มค้นหาคนเดียว โดยที่พวกเขารู้ว่าเป้าหมายของกลุ่มต้องการค้นหาอะไร แต่จะไม่มีผลการค้นหาของเพื่อนในทีมให้พวกเขา
- Continue Scenario คือผู้ทดลองจะเริ่มค้นหาข้อมูลต่อจากผู้ค้นหาก่อนหน้า โดยที่พวกเขารู้ว่าเป้าหมายของกลุ่มต้องการค้นหาอะไร และมีผลการค้นหาของเพื่อนในทีมให้พวกเขารับรู้
- Search Scenario คือผู้ทดลองจะเริ่มค้นหาข้อมูลต่อจากผู้ค้นหาก่อนหน้า โดยที่พวกเขาไม่รู้ว่าเป้าหมายของกลุ่มต้องการค้นหาอะไร และมีผลการค้นหาของเพื่อนในทีมให้พวกเขารับรู้ พวกเขาต้องพยายามทำความเข้าใจเป้าหมายของกลุ่มจากผลการค้นหาก่อนหน้าที่เพื่อนค้นหาไว้ให้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการทดลองพวกเขาพบว่า Continue Scenario และ Search Scenario ให้ผลการค้นหาที่ไม่แตกต่างกัน แสดงว่าการเครื่องมือของ ViGOR ทำให้ผู้ค้นหาสามารถรับรู้ได้ว่าเป้าหมายของกลุ่มต้องการค้นหาอะไร แม้ไม่ได้บอกพวกเขาว่าต้องค้นหาเกี่ยวกับอะไร แต่พวกเขารับรู้ได้จากผลการค้นหาของเพื่อนร่วมทีมก่อนหน้า และเมื่อเปรียบเทียบการค้นหาคนเดียวกับการค้นหาแบบร่วมมือก็พบความแตกต่างเพียงเล็กน้อยเท่านั้น นอกจากนี้พวกเขายังพบว่าเครื่องมือของ ViGOR ไม่มีผลกระทบเชิงลบต่อการค้นหาของผู้ใช้ ผู้ใช้สามารถค้นหาวีดีโอแบบร่วมมือได้สะดวกมากขึ้นกว่าแบบเก่าที่ไม่มี ViGOR



รูปที่ 2.3 ตัวอย่างระบบค้นหาข้อมูล ViGOR

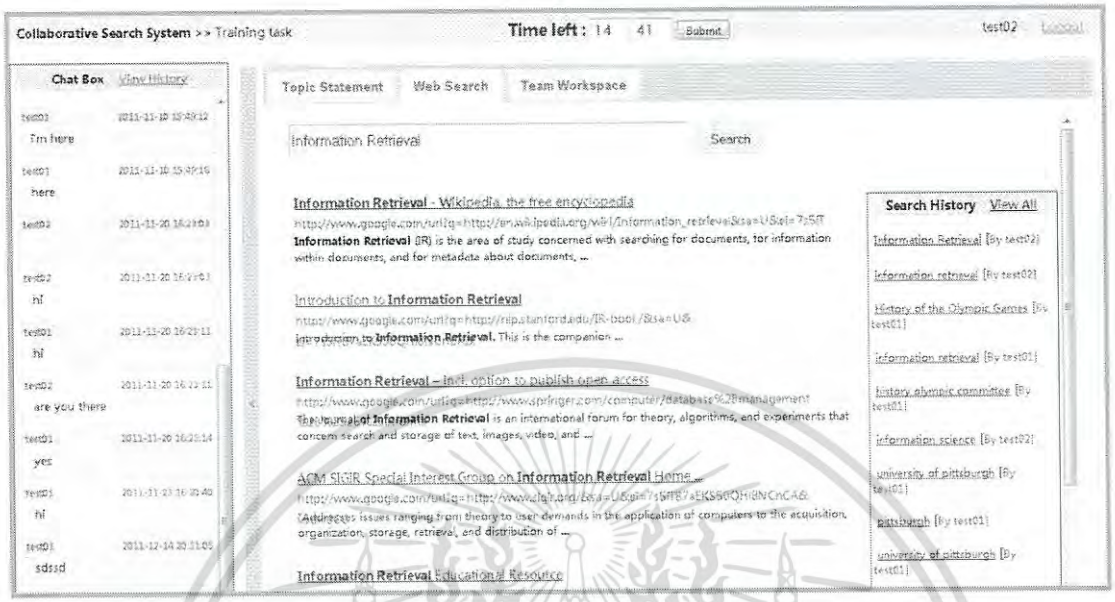
3. CollabSearch

CollabSearch ถูกพัฒนาโดย Z. Yue และคณะ (2012) ซึ่งเป็นระบบค้นหาข้อมูลที่ถูกออกแบบเพื่อการค้นหาข้อมูลร่วมกันทั้งในเวลาเดียวกันและต่างเวลากัน [15] รูปที่ 2.4 แสดงตัวอย่างของระบบ CollabSearch โดยระบบมีประกอบด้วย 3 หน้าต่าง ดังนี้คือ

- Topic Statement เพื่อให้ผู้ค้นหาสามารถบอกเพื่อนในทีมให้รับรู้ว่า พวกเขากำลังค้นหาข้อมูลเกี่ยวกับเรื่องอะไรหรืออาจจะเป็นความคิดเห็นของผู้ค้นหา
- Web Search เป็นหน้าต่างสำหรับใช้ค้นหาและแสดงผลการค้นหาเมื่อผู้ใช้ใส่คำค้นหา โดย CollabSearch ใช้ผลการค้นหาของ google นอกจากนี้ยังมีเครื่องมือในการบันทึกเว็บไซต์ที่เกี่ยวข้องและแสดงความคิดเห็นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Team Workspace เป็นหน้าต่างแสดงผลข้อมูลที่ผู้ค้นหาได้บันทึกไว้ระหว่างค้นหาข้อมูลจากหน้า Web Search



รูปที่ 2.4 ตัวอย่างระบบค้นหาข้อมูล CollabSearch

ในการทดลอง พวกเขาอธิบายถึงการค้นหาข้อมูลแบบร่วมกันที่มีหลายรูปแบบ และพฤติกรรมของผู้ใช้งานในด้านการทำงานร่วมกัน โดยศึกษาจากรูปแบบการทำงานที่แตกต่างกันและนำเสนอการเปรียบเทียบความแตกต่างของวิธีการค้นหาข้อมูล 3 แบบ โดยมีดังนี้คือ

- ผู้ใช้ค้นหาข้อมูลร่วมกันในเวลาเดียวกัน แบบมีการสื่อสารที่ชัดเจน (Synchronously with explicit communication or CEC)
- ผู้ใช้ค้นหาข้อมูลร่วมกันแบบต่างเวลา แบบไม่มีการสื่อสารที่ชัดเจน (Asynchronously without explicit communication or COC)
- ผู้ใช้ค้นหาข้อมูลคนเดียว ไม่สื่อสารกัน (Individual search or IND)

ตารางที่ 2.1 แสดงข้อมูลที่พวกเขาเก็บเพื่อการวิเคราะห์ผลทดลอง ซึ่งเขาแบ่งกิจกรรมของผู้ใช้ออกเป็น 6 รูปแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 กิจกรรมที่ผู้ทดลองกระทำระหว่างค้นหาข้อมูล

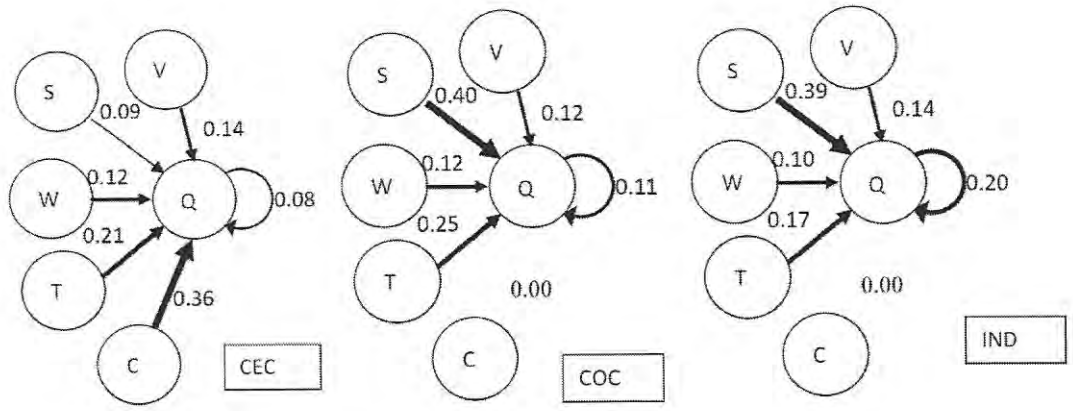
Action	Description
Query (Q)	จำนวนคำที่ผู้ค้นหาใช้ในการค้นหาทั้งหมด
View (V)	จำนวนเว็บไซต์ทั้งหมดที่ผู้ค้นหาเข้าชม
Save (S)	จำนวนเว็บไซต์ที่เกี่ยวข้องที่ผู้ใช้ทำการบันทึก
Workspace (W)	จำนวนความคิดเห็นที่ผู้ค้นหาบันทึกลงใน Workspace
Topic (T)	จำนวนครั้งที่ผู้ค้นหาเข้าไปดูเป้าหมายของงานที่ต้องการค้นหา
Chat (C)	จำนวนบทสนทนาที่ผู้ค้นหาสื่อสารกันระหว่างทำการค้นหา

ในการวิเคราะห์ข้อมูล พวกเขาคิดว่าการดูจำนวนการกระทำของผู้ค้นหามาเปรียบเทียบนั้นไม่เพียงพอ พวกเขาสนใจจำนวนของการเปลี่ยนแปลงการกระทำของผู้ใช้ ทำให้ได้การเปลี่ยนแปลงการกระทำระหว่างการค้นหาข้อมูลทั้งหมด 36 รูปแบบ ดังรูปภาพที่ 2.5

รูปที่ 2.5 เปอร์เซ็นต์ของคู่การกระทำใน 3 เงื่อนไข

ผลการทดลองแสดงในรูปที่ 2.6 พวกเขามองว่าการสร้างคำที่ใช้ในการค้นหา (Query) เป็นสิ่งที่สำคัญในเรื่องของการค้นหา พวกเขาจึงสังเกตว่ากิจกรรมใดที่มีผลต่อการสร้างคำที่ใช้ในการค้นหา ผลการทดลองพบว่า ในการค้นหาข้อมูลแบบ CEC ฟังก์ชันการสนทนา (Chat) มีอิทธิพลต่อการสร้างคำที่ใช้ค้นหามากที่สุด ในส่วนของการค้นหาแบบ COC และ IND ฟังก์ชันบันทึกเว็บไซต์ (Save) มีอิทธิพลต่อการสร้างคำที่ใช้ในการค้นหามากที่สุด จึงสรุปได้ว่าฟังก์ชันสนทนาเป็นฟังก์ชันที่สำคัญในการค้นหาข้อมูลร่วมกันแบบ CEC ส่วนฟังก์ชั่นการบันทึกเว็บไซต์เป็นฟังก์ชันที่สำคัญในการค้นหาข้อมูลร่วมกันแบบ COC และ IND

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



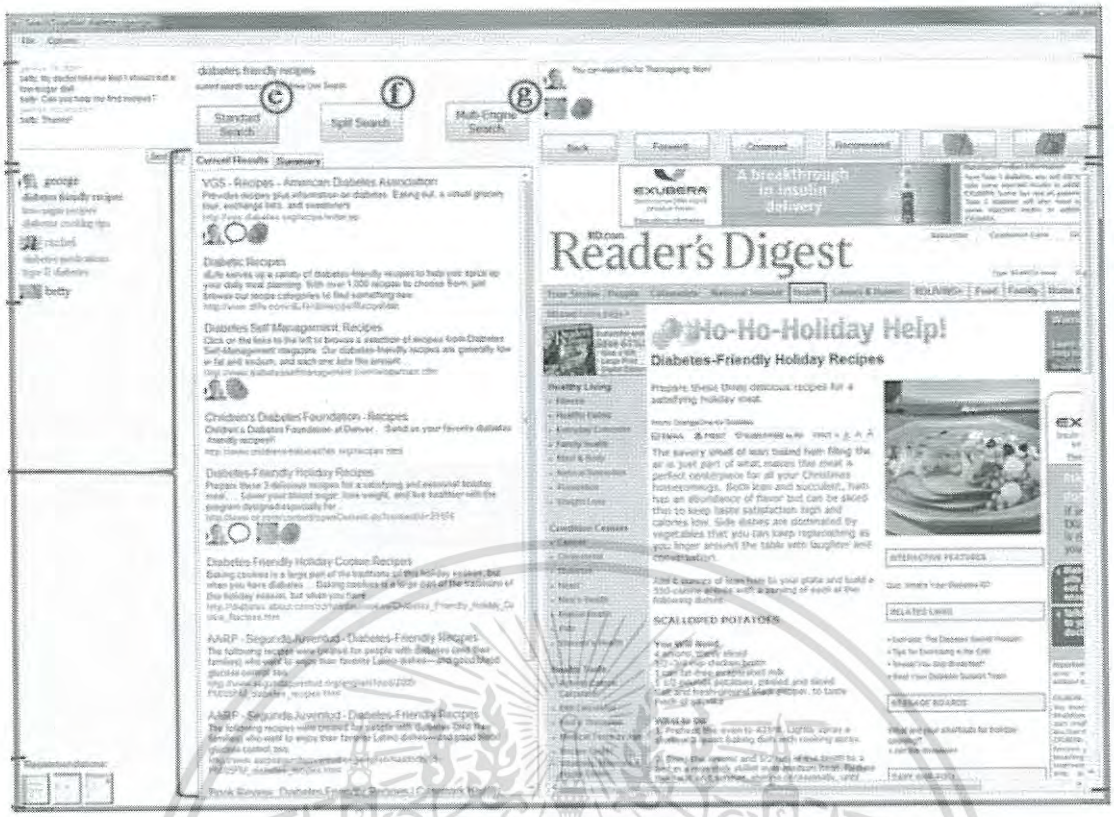
รูปที่ 2.6 แสดงปริมาณกิจกรรมที่ผู้ทดลองทำก่อนการใส่คำค้นหาข้อมูลของทั้ง 3 รูปแบบการค้นหา

4. SearchTogether

SearchTogether ถูกพัฒนาโดย M. R. Morris และ E. Horvitz (2007) จาก Microsoft โดยพวกเขาออกแบบ SearchTogether ตามงานวิจัยของพวกเขาจากการสำรวจ ตัวอย่างระบบ SearchTogether แสดงในรูปที่ 2.7 พวกเขาพบว่า ระบบค้นหาข้อมูลแบบร่วมมือกันที่ดี ควรจะออกแบบเพื่อสนับสนุนการทำงานร่วมกัน [16] ซึ่งประกอบด้วย 3 องค์ประกอบนี้คือ 1) การรับรู้ 2) การแบ่งงาน และ 3) การคงอยู่ ทำให้ระบบ SearchTogether ถูกออกแบบตามองค์ประกอบดังกล่าว ซึ่งจะอธิบายดังต่อไปนี้คือ

- การรับรู้ (Awareness) ซึ่งในระบบ SearchTogether มีการแสดงสถานะของผู้ใช้ถ้ารูปของผู้ใช้เป็นสีเทาแสดงว่า พวกเขาไม่อยู่ในระบบแล้ว ในการค้นหาข้อมูลจะมีการบันทึกข้อมูลการค้นหา เพื่อลดความซ้ำซ้อนการใช้คำในการค้นหาเดิม เข้าเว็บไซต์เดิม และมีเครื่องมือในการระบุเว็บไซต์ที่เกี่ยวข้องและแบ่งปันให้เพื่อนที่ค้นหาร่วมกันอีกด้วย
- การแบ่งงาน (Division of Labor) ซึ่งในระบบ SearchTogether เมื่อผู้ใช้พบเว็บไซต์ที่ต้องการให้เพื่อนในทีมเห็น พวกเขาสามารถเลือกเพื่อนในทีมที่ต้องการแนะนำเว็บไซต์นั้นได้ หรือแสดงความคิดเห็นเว็บไซต์ นอกจากนี้ระบบยังมีเครื่องมือในการสนทนาแบบทันที (instant messaging)
- การคงอยู่ (Persistence) ในระบบ SearchTogether เมื่อผู้ใช้เข้าสู่ระบบเขาสามารถเห็นข้อมูลต่างๆ ที่ผู้ใช้คนอื่นได้ทำไว้ในตอนที่เขาออฟไลน์และเก็บข้อมูลต่างๆ ที่เขาทำไว้ล่าสุด เพื่อเป็นการเตือนให้รู้ว่าทำอะไรไปแล้วบ้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 ตัวอย่างระบบค้นหาข้อมูล SearchTogether

5. WeSearch

WeSearch ถูกพัฒนาโดย M. R. Morris และคณะ (2010) ซึ่งเป็นระบบค้นหาข้อมูลแบบร่วมมือกันที่ประยุกต์โต๊ะมาใช้ในการแสดงผล (table-top) ในรูปที่ 2.8 แสดงตัวอย่างการค้นหาข้อมูลร่วมกันบนโต๊ะขนาด 4x6 ซึ่งรองรับการทำงานได้เพียง 4 คนเท่านั้น และเป็นการทำงานแบบในเวลาเดียว และสถานที่เดียวกันเท่านั้น พวกเขาได้ทำการศึกษาว่างานประเภทไหนที่เหมาะกับระบบ WeSearch ในการทดลองพวกเขามีโอกาสได้สังเกตกลุ่มผู้ใช้หลายประเภท ได้แก่ ครอบครัว นักเรียน และเพื่อนร่วมงาน ในการค้นหาข้อมูลร่วมกัน และพบว่าการค้นหาข้อมูลบนโต๊ะ โดยใช้ระบบ WeSearch เหมาะกับสถานการณ์ที่ผู้เชี่ยวชาญมืออาชีพและลูกค้าทำงานร่วมกัน เช่น บรรณารักษ์และนักเรียนหาข้อมูลร่วมกัน หรือพนักงานขายช่วยลูกค้าในการค้นหาผลิตภัณฑ์ที่มีความหลากลายด้วยกัน เป็นต้น [17]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.8 ตัวอย่างระบบค้นหาข้อมูล Wesearch

6. Cofox

Cofox ถูกพัฒนาโดย J. R. Perez และคณะ (2012) ซึ่งเป็นระบบค้นหาข้อมูลที่ถูกออกแบบเพื่อการค้นหาข้อมูลร่วมกันในเวลาเดียวกันเท่านั้น โดยผู้ค้นหาจะต้องทำงานร่วมกันเพียงแค่ 2 คน โดยระบบจะทำการแสดงภาพถ่ายวิดีโอแบบเรียลไทม์ของกิจกรรมที่ผู้ใช้ทำระหว่างค้นหาให้ผู้ค้นหาอีกคนรับรู้ราวกับว่าพวกเขาทำงานร่วมกันในที่เดียวกัน [18] ตัวอย่างระบบค้นหาข้อมูล Cofox แสดงในรูปที่ 2.9 ซึ่งในระบบจะมี 2 หน้าต่าง ดังนี้คือ

- หน้าต่างสำหรับค้นหาข้อมูลของผู้ใช้ (The Local User Area) ซึ่งอยู่ทางด้านขวา เป็นหน้าต่างเพื่อให้ผู้ค้นหาสามารถใช้ในการค้นหาข้อมูล
- หน้าต่างแสดงการทำงานของผู้ร่วมค้นหาระยะไกล (The Remote User Area) ซึ่งอยู่ทางด้ายซ้าย เป็นหน้าต่างที่ทำให้ผู้ค้นหาสามารถรับรู้ว่าผู้ร่วมค้นหาระยะไกลเข้าเว็บไซต์ไหนอยู่ หรือใช้คำค้นหาอะไร โดยการแสดงวิดีโอภาพแบบเรียลไทม์ นอกจากนี้ยังมีเครื่องมือในการสนทนา (instant messaging) เพื่อปรึกษากันในการค้นหาข้อมูล

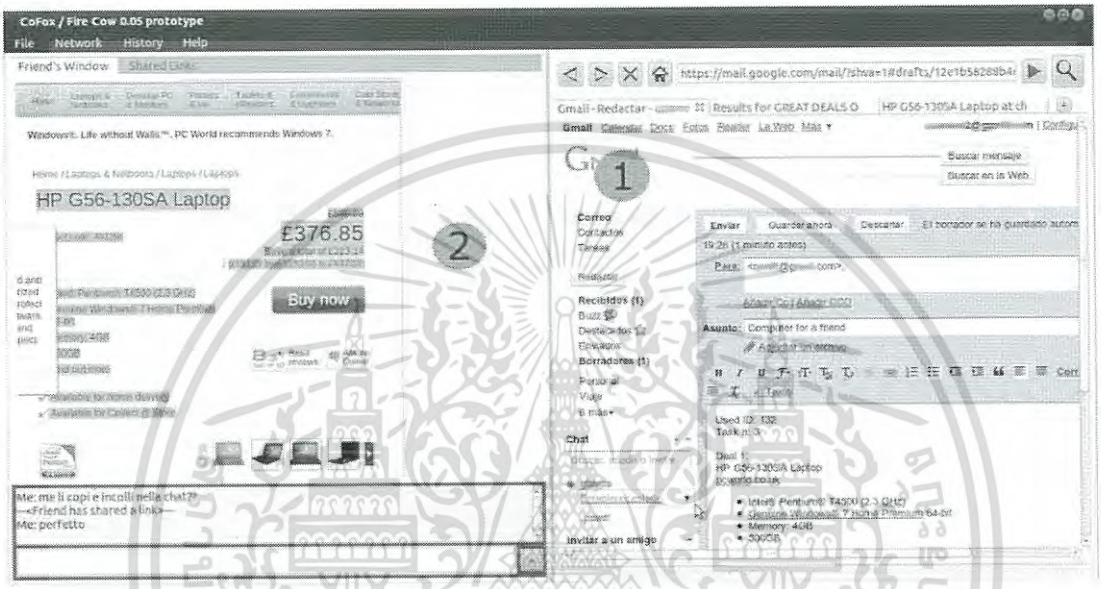
ในการทดลอง พวกเขาทำการวัดประสิทธิภาพการทำงานของระบบ Cofox ในการค้นหาข้อมูลโดยทำการเปรียบเทียบกับระบบอื่นอีก 2 ระบบคือ

- ระบบ Solo คือระบบที่รองรับการค้นหาข้อมูลเพียงคนเดียว
- ระบบ Tog คือระบบที่รองรับการค้นหาข้อมูล 2 คน แต่ใช้คอมพิวเตอร์ 1 เครื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวิเคราะห์ข้อมูล พวกเขาเก็บข้อมูลทั้งในเชิงปริมาณและคุณภาพ โดยข้อมูลในเชิงคุณภาพพวกเขาเก็บจากแบบสอบถามหลังการทดลอง ในส่วนของข้อมูลเชิงปริมาณ พวกเขาได้ทำการเก็บข้อมูลระหว่างการค้นหาของผู้ทดลองเพื่อนำมาเปรียบเทียบ ซึ่งมีดังนี้คือ

- จำนวนเว็บไซต์ที่เกี่ยวข้องที่ผู้ค้นหาระบุ
- จำนวนเว็บไซต์ที่ผู้ทดลองเข้าไปดูทั้งหมด
- เวลาที่ทำการค้นหาข้อมูลจากงานที่ได้รับ



รูปที่ 2.9 ตัวอย่างระบบค้นหาข้อมูล Cofox

ผลการทดลองแสดงในตารางที่ 2.1 พบว่าระบบ Cofox มีประสิทธิภาพดีว่าระบบ Solo และ Tog เพราะผู้ใช้ได้รับเว็บไซต์ที่ต้องการมากกว่า และเมื่อเปรียบเทียบจำนวนเว็บไซต์ที่ตรงความต้องการกับจำนวนเว็บไซต์ที่เข้าดูทั้งหมด เราพบว่า Cofox มีโอกาสเจอเว็บไซต์ที่ตรงความต้องการมากกว่าคิดเป็น 23.92%

ตารางที่ 2.2 ข้อมูลเชิงปริมาณเปรียบเทียบระบบ Cofox กับระบบ Solo และ Tog

System	Rel. Pages	Tot. Pages	Time	RelPerMin	%
Solo	257	1567	19	13.00	15.76
Cofox	**393**	**1643**	**17**	**23.12**	**23.92**
Tog	126	744	12	10.50	16.94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 งานวิจัยที่เกี่ยวข้องกับการสรุปแบบภาพ (Visual Snippet)

1. Visual Snippets: Summarizing Web Pages for Search and Revisitation

J. Teevan และคณะ (2009) ได้ค้นหาวิธีการแสดงเว็บไซต์แทนการเก็บในรูปแบบข้อความตัวอักษร URL ของเว็บไซต์ หลังจากที่เว็บไซต์นั้นได้ถูกระบุว่าเกี่ยวข้องกับความต้องการ พวกเขาพบความแตกต่างของแต่ละเว็บไซต์ 3 อย่างคือ 1) รูปภาพที่สำคัญ 2) โลโก้ 3) ชื่อเรื่อง และทำการดึงข้อมูลทั้ง 3 อย่างมารวมกันเป็นรูปภาพสรุปของเว็บไซต์นั้น เพื่อให้ผู้ค้นหารู้สึกง่ายในการค้นหาหรือการกลับไปดูเว็บไซต์นั้นอีกครั้ง ผลการศึกษาของพวกเขาพบว่าการแสดงเว็บไซต์แบบรูปภาพสรุปไม่แตกต่างจากข้อความตัวอักษรในเรื่องการค้นหาเว็บไซต์ใหม่ แต่ในเรื่องการกลับมาดูเว็บไซต์ที่เคยเข้าดูแล้วอีกครั้ง สรุปแบบภาพทำให้ผู้ใช้จำเว็บไซต์นั้นได้ดีว่าแบบข้อความตัวอักษร [19] ตัวอย่างสรุปแบบภาพแสดงในรูปที่ 2.10

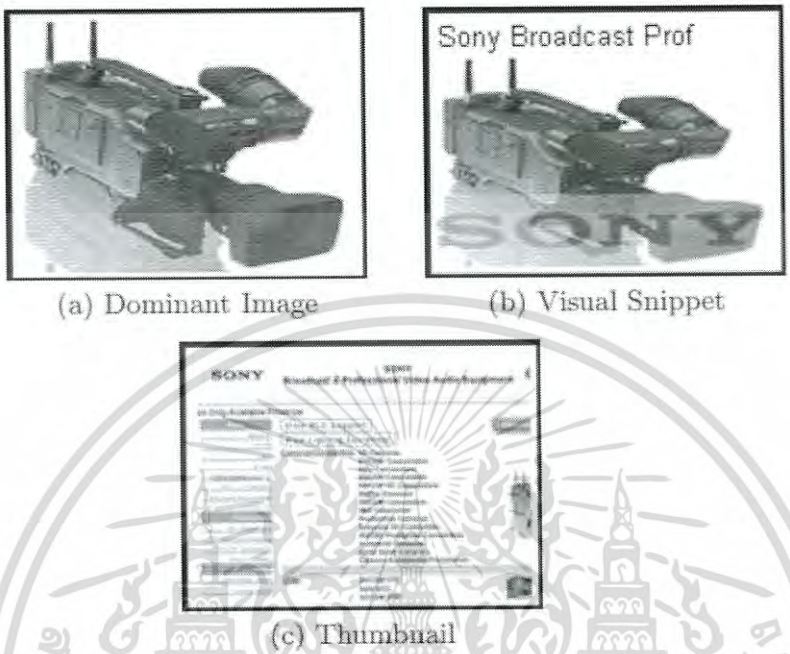
รูปที่ 2.10 (ล่าง) ตัวอย่างเว็บไซต์ และ (บน) สรุปแบบภาพ

2. Visual Summarization of Web Pages

B. Jiao และคณะ (2010) ได้ทำการวิเคราะห์การทำสรุปแบบภาพ ของ J. Teevan และพวกเขาพบปัญญาคือ มีเว็บไซต์มากมายที่ไม่สามารถทำสรุปแบบภาพ โดยวิธีของ J. Teevan ได้ เพราะในเว็บไซต์เหล่านั้นไม่มีรูปภาพ (internal image) ที่จะเอาไปใช้เป็นส่วนหนึ่งของการทำสรุปแบบภาพ พวกเขานำเสนออีกวิธีการในการทำสรุปแบบภาพ คือ การหารูปภาพจากภายนอกเว็บไซต์ (external image) ที่เกี่ยวข้องกับเว็บไซต์นั้น เพื่อช่วยในการสร้างสรุปแบบภาพจากเว็บไซต์ที่ไม่มีรูปภาพ ตัวอย่างการทำสรุปแบบภาพ แสดงใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.11 ผลการตรวจสอบประสิทธิภาพพบว่า รูปภาพภายในเว็บไซต์ให้ความถูกต้องน่าเชื่อถือในการสรุปเว็บไซต์ แต่รูปภาพนอกเว็บไซต์มีประโยชน์ในการช่วยส่งเสริมและเติมเต็มรูปภาพภายในที่มีอยู่ในการสรุปเว็บไซต์ได้มากขึ้น [20]



รูปที่ 2.11 ตัวอย่าง (a) รูปภาพที่โดดเด่นที่ดึงได้จากเว็บไซต์ (b) สรุปแบบภาพ และ (c) รูปภาพขนาดย่อของเว็บไซต์จริง

3. A Comparison of Visual and Textual Page Previews in Judging the Helpfulness of Web Pages

A. Aula และคณะ (2010) ได้แนะนำการใช้รูปภาพในการนำเสนอเว็บไซต์แทนข้อความตัวอักษร เช่น title หรือ URL ของเว็บไซต์ แต่รูปภาพของ A. Aula ต่างจากรูปภาพของ J. Teevan ที่เป็นรูปภาพเว็บไซต์ขนาดย่อ (thumbnail) ผลการทดลองการเปรียบเทียบข้อความตัวอักษรกับรูปภาพในการนำเสนอเว็บไซต์พบว่า ข้อความตัวอักษรมีแนวโน้มทำให้ผู้ค้นหาประเมินค่าเว็บไซต์สูงเกินไปทำให้ผู้ค้นหาเข้าไปดูเว็บไซต์นั้นแต่ไม่ตรงกับความต้องการ และสรุปแบบภาพมีแนวโน้มทำให้ผู้ค้นหาประเมินค่าเว็บไซต์ต่ำเกินไปทำให้ผู้ค้นหาไม่เข้าไปดูเว็บไซต์นั้น พวกเขาแนะนำว่าการรวมกันของรูปภาพและข้อความตัวอักษรจะทำให้มีประสิทธิภาพมากที่สุดในการใช้เป็นตัวแทนนำเสนอเว็บไซต์ รูปที่ 2.12 แสดงตัวอย่างการรวมกันของรูปภาพเว็บไซต์ขนาดย่อและข้อความตัวอักษร [21]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.12 ตัวอย่างการนำรูปภาพขนาดย่อมารวมกับสรุปข้อความตัวอักษรของเว็บไซต์

2.2.3 งานวิจัยที่เกี่ยวข้องกับคำแนะนำสำหรับการค้นหา (Query Suggestion)

คำแนะนำสำหรับการค้นหาส่วนใหญ่ในระบบค้นหาข้อมูลถูกนำเสนอในรูปแบบข้อความตัวอักษร ซึ่งยังมีปัญหาเรื่องช่องว่างของความตั้งใจ (intention gap) ระหว่างความตั้งใจของผู้ค้นหา (intention) และคำที่พวกเขาใช้ค้นหา (query) เพื่อที่จะลดช่องว่างนี้ จึงมีนักวิจัยคิดวิธีการนำรูปภาพมาช่วยในการแนะนำ ดังนี้คือ

1. Visual Query Suggestion (VQS)

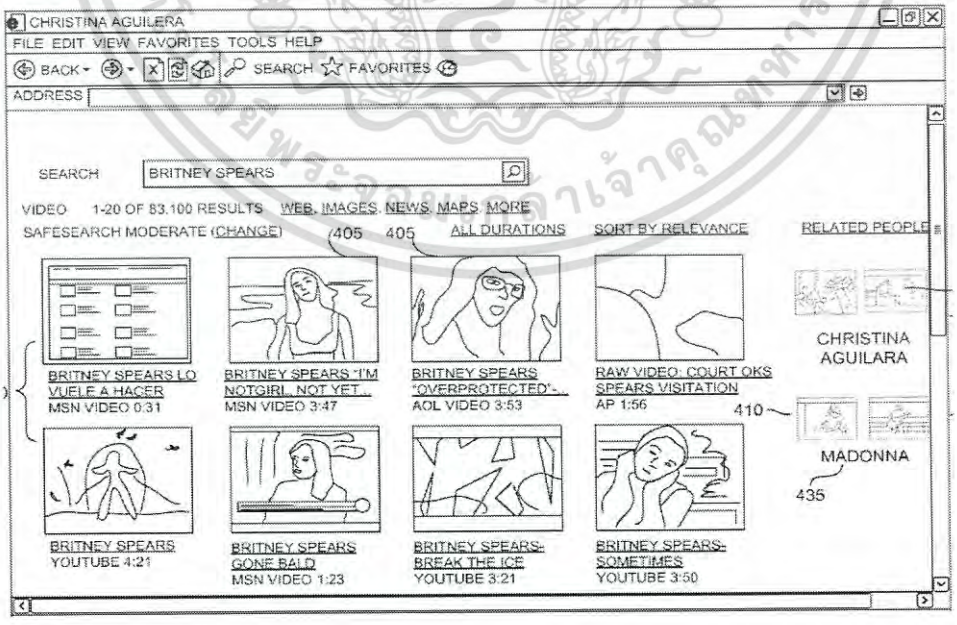
Z. J. Zha และคณะ (2009) นำเสนอการใช้รูปภาพเพิ่มเติมในการแนะนำคำเพื่อการค้นหาข้อมูล แทนการแนะนำจากตัวอักษรอย่างเดียว งานวิจัยของพวกเขาพบว่า รูปภาพช่วยให้ผู้ค้นหาใช้คำหรือเลือกคำในการค้นหาได้แม่นยำมากขึ้นและทำให้ได้ข้อมูลที่ตรงกับความต้องการรวดเร็วขึ้น รูปที่ 2.13 แสดงตัวอย่างการแนะนำรูปภาพเพื่อการค้นหาข้อมูล [22]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.13 ตัวอย่างการแนะนำรูปภาพเพื่อการค้นหาข้อมูล

2. Video Query Suggestion

J. Denney และคณะ (2008) ได้นำเสนอวิธีการแนะนำคำที่เกี่ยวข้องสำหรับการค้นหาข้อมูลประเภทวีดีโอ เมื่อผู้ใช้ใส่คำค้นหาเพื่อค้นหาข้อมูลประเภทวีดีโอ ระบบของพวกเขาไม่เพียงแค่ให้ข้อมูลประเภทวีดีโอออกมา แต่ยังแนะนำวิดีโอที่น่าจะเกี่ยวข้องกับคำที่ผู้ค้นหาใช้อีกด้วย รูปที่ 2.14 ตัวอย่างการแนะนำวีดีโอเพื่อการค้นหาข้อมูล [23]

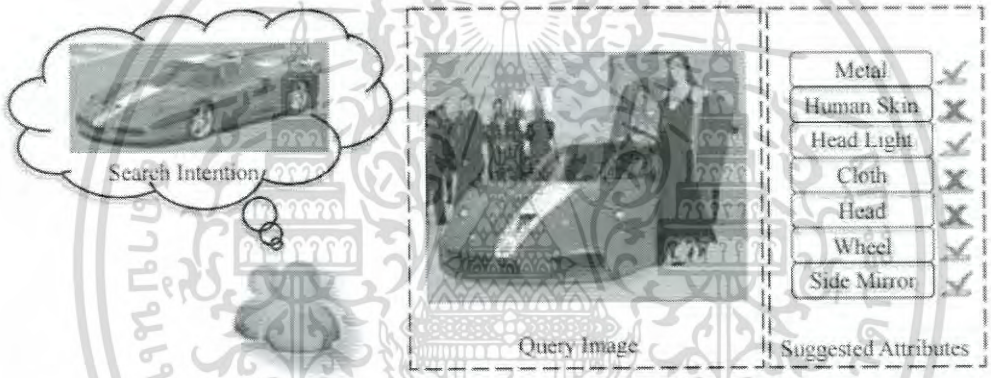


รูปที่ 2.14 ตัวอย่างการแนะนำวีดีโอเพื่อการค้นหาข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. Visual Query Attributes Suggestion (VQAS)

J. Bian และคณะ (2012) ได้นำเสนอการปรับปรุงและพัฒนาการแนะนำรูปภาพเพื่อการค้นหาข้อมูล (VQS) ของจาก Z. J. Zha โดยพวกเขาพบปัญหา 2 ข้อคือ 1) รูปภาพที่แนะนำเพื่อการค้นหามีองค์ประกอบอื่นบางส่วนที่ไม่เกี่ยวข้องหรือเป็นประโยชน์ต่อข้อมูลที่ต้องการ ซึ่งทำให้ผู้ใช้สับสนในการเลือกภาพแนะนำในการค้นหา 2) รูปภาพที่แนะนำเพื่อการค้นหาไม่มีองค์ประกอบที่เกี่ยวข้องเลย พวกเขาจึงแนะนำวิธีการเพิ่มเติม คือ เมื่อผู้ค้นหาเลือกรูปภาพที่ใช้ในการค้นหาแล้ว ระบบจะมีการประมวลผลหาองค์ประกอบหรือคุณลักษณะภายในรูปภาพ (attributes) และนำคุณลักษณะภายในรูปภาพมาแนะนำให้ผู้ค้นหาเพิ่มเติม ซึ่งทำให้ผู้ค้นหาสามารถระบุความต้องการได้แม่นยำมากขึ้น รูปที่ 2.15 ตัวอย่างการแนะนำรูปภาพและองค์ประกอบภายในรูปภาพเพื่อการค้นหาข้อมูล [24]



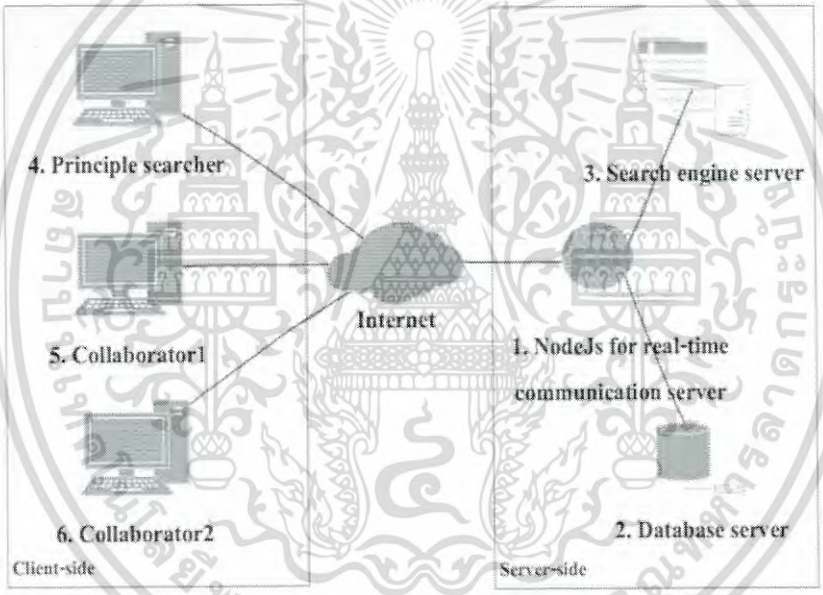
รูปที่ 2.15 ตัวอย่างการแนะนำรูปภาพและองค์ประกอบภายในรูปภาพเพื่อการค้นหาข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
โคสเปช: ระบบค้นหาข้อมูลแบบร่วมมือ

ระบบค้นหาข้อมูลแบบร่วมมือในปัจจุบันมีการนำรูปภาพมาใช้ช่วยในการสื่อสารระหว่างค้นหาข้อมูลมากขึ้น แต่ยังไม่เป็นที่นิยมและต้องค้นหาข้อมูลพร้อมกันในเวลาเดียวกันเท่านั้น เราจึงพัฒนาระบบโคสเปชขึ้นเพื่อแก้ปัญหานี้ บทนี้นำเสนอโครงสร้างและรายละเอียดของระบบโคสเปช

3.1 สถาปัตยกรรมของระบบโคสเปช



รูปที่ 3.1 ตัวอย่างสถาปัตยกรรมของระบบโคสเปช

3.1.1 Node.js Real-Time Communication Server

เซิร์ฟเวอร์ที่ให้บริการการติดต่อสื่อสารแบบเรียลไทม์ ตัวอย่างเช่น เมื่อผู้ค้นหาทำการค้นหา จะมีการกระจายแบบทุกทิศทาง (broadcast) คำที่ใช้ค้นหา ให้ผู้ที่ร่วมค้นหา (collaborator) ได้รู้ว่า ผู้ค้นหาค้นหาด้วยคำอะไร หรือที่เห็นได้ชัดเจนคือ การสนทนาระหว่างค้นหา (chat)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 Database Server

เซิร์ฟเวอร์ที่รวบรวมข้อมูลต่าง ๆ ที่เกี่ยวข้องกันเข้าไว้ด้วยกันอย่างมีระบบมีความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ที่ชัดเจน ในระบบฐานข้อมูลจะประกอบด้วยแฟ้มข้อมูลหลายแฟ้ม มีการเก็บข้อมูลที่เกี่ยวข้องและสัมพันธ์กันเข้าไว้ด้วยกันอย่างเป็นระบบ ตัวอย่างเช่น ในระบบโค-สเปช ก่อนที่ real-time communication server จะทำการกระจายข้อมูลผลการค้นหาหรือกิจกรรมต่าง ๆ ให้แก่ ผู้ร่วมค้นหาอื่นได้รับรู้ real-time communication server จะส่งข้อมูลผลการค้นหาหรือกิจกรรมต่าง ๆ ที่เกิดขึ้นไปบันทึกไว้ใน database server ก่อน

3.1.3 Search Engine Server

เซิร์ฟเวอร์ที่ใช้ในการรวบรวมดัชนีของเว็บไซต์ ช่วยในการค้นคืนข้อมูลโดยเฉพาะข้อมูลบนอินเทอร์เน็ต โดยครอบคลุมทั้งข้อความ รูปภาพ ภาพเคลื่อนไหว เพลง ซอฟต์แวร์ แผนที่ ข้อมูลบุคคล กลุ่มข่าว และอื่น ๆ ซึ่งแตกต่างกันไปแล้วแต่โปรแกรมหรือผู้ให้บริการแต่ละราย ระบบค้นหาส่วนใหญ่จะค้นหาข้อมูลจากคำสำคัญที่ผู้ค้นหาใส่เข้าไป จากนั้นก็จะแสดงรายการผลลัพธ์ที่คิดว่าผู้ค้นหาน่าจะต้องการ โดยในปัจจุบันระบบค้นหาบางตัว เช่น google จะมีการบันทึกประวัติการค้นหา และการเลือกผลการค้นหาของผู้ใช้ เพื่อนำประวัติที่บันทึกไว้ มาปรับปรุงผลการค้นหาของระบบในการค้นหาครั้งต่อไป

3.1.4 Principle Searcher

บุคคลที่ทำการสร้างโปรเจคเพื่อทำการค้นหาข้อมูล พวกเขาสามารถชวนเพื่อนเข้ามาในโปรเจค เพื่อค้นหาข้อมูลร่วมกันในระบบ และเป็นเพียงคนเดียวที่มีสิทธิ์ในการปิดโปรเจคที่ตัวเองสร้างขึ้น

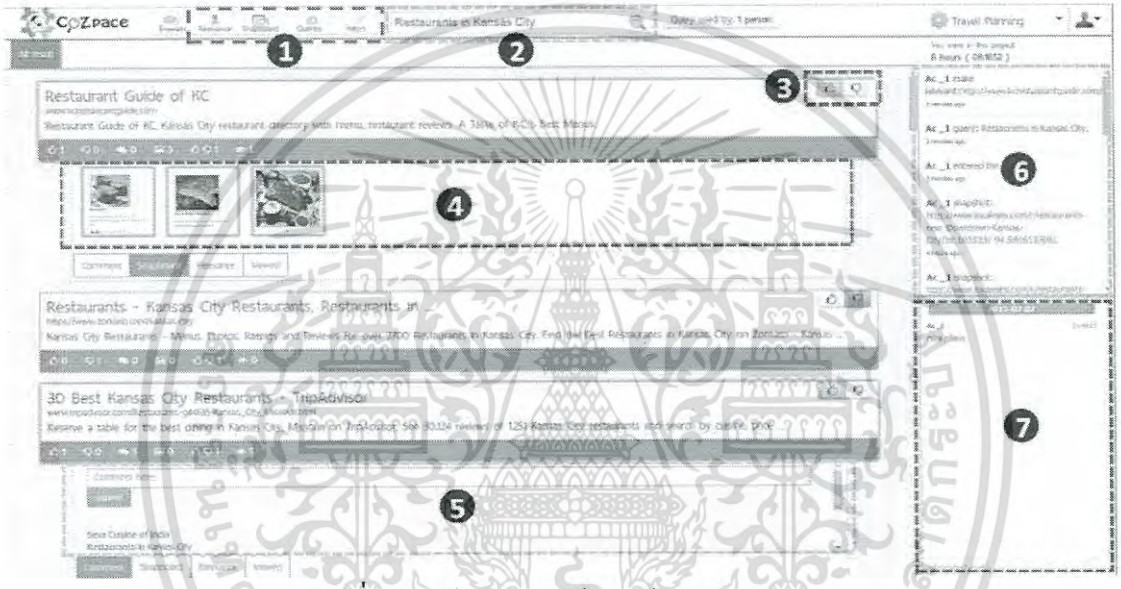
3.1.5 Collaborators

บุคคลที่ถูก principle searcher เพิ่มเข้ามาในโปรเจค เพื่อช่วยในการค้นหาข้อมูลร่วมกัน ซึ่งสามารถทำกิจกรรมต่าง ๆ ที่ระบบมีให้ได้เหมือนกับ principle searcher แตกต่างกันตรงที่ collaborators ไม่มีสิทธิ์ในการปิดโปรเจค แต่สามารถที่จะออกจากโปรเจค เมื่อไม่ต้องการที่จะเข้าร่วมการค้นหานั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 โคสเปช (CoZpace)

โคสเปช [2][3] เป็นเว็บแอพพลิเคชัน ซึ่งถูกออกแบบมาเพื่อช่วยให้ผู้ใช้ค้นหาข้อมูลร่วมกันได้อย่างมีประสิทธิภาพมากขึ้น ระบบนี้มีเครื่องมือช่วยในการติดต่อสื่อสารร่วมกันระหว่างผู้ค้นหาในกลุ่มระหว่างที่ทำการค้นหาข้อมูล เช่น การสนทนาร่วมกัน การดูประวัติคำค้นหา การแสดงความคิดเห็นสำหรับเว็บไซต์ที่ถูกค้นคืนออกมา เป็นต้น รูปที่ 3.1 นำเสนอส่วนติดต่อผู้ใช้ของระบบโคสเปช ซึ่งแสดงผลการค้นหาหลังใส่คำที่ใช้ในการค้นหา (query) ตัวอย่างคำที่ใช้ในการค้นหาคือ "ร้านอาหารในเมืองแคนซัส" (Restaurant in Kansas City) และส่วนประกอบทั้ง 7 ส่วนในรูปที่ 3.1 จะอธิบายดังต่อไปนี้

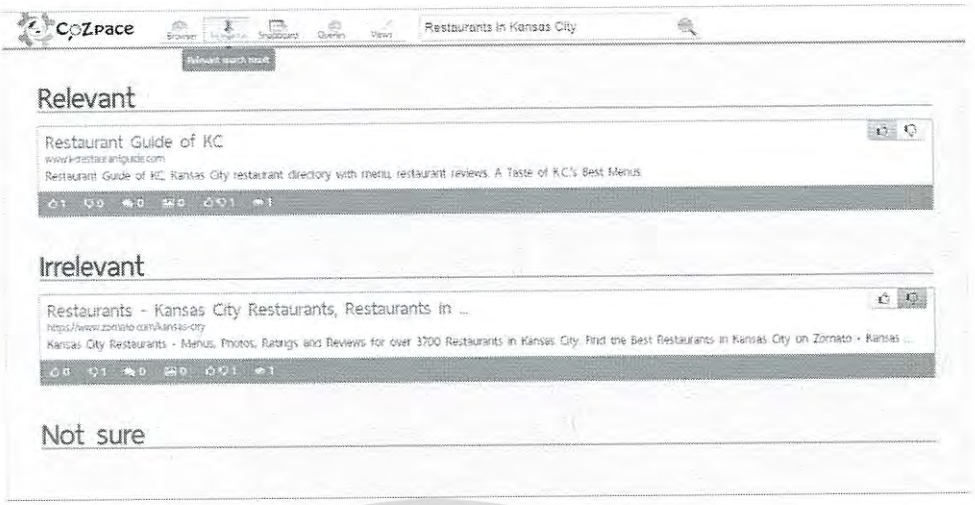


รูปที่ 3.2 หน้าตาระบบค้นหาข้อมูล CoZpace

1. แถบสรุปผลการค้นหาประกอบด้วย 4 ปุ่ม ดังนี้คือ

   1.1) ปุ่ม Relevance แสดงเว็บไซต์แบ่งออกเป็น 3 กลุ่ม ตามการตัดสินของผู้ค้นหาภายในกลุ่ม ได้แก่ เว็บไซต์ที่เกี่ยวข้อง (relevant website) เว็บไซต์ที่ไม่เกี่ยวข้อง (non-relevant website) และเว็บไซต์ที่ยังไม่แน่ใจ (not sure)

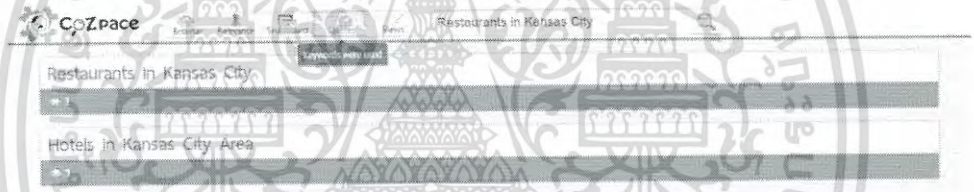
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 หน้าต่างแสดงผลหลังกดปุ่ม Relevance

1.2) ปุ่ม Snapboard แสดงรูปภาพทั้งหมดที่ผู้ใช้ตัด จะแสดงตัวอย่างในรูปที่ 3.6

1.3) ปุ่ม Queries แสดงคำที่ใช้ค้นหาทั้งหมด



รูปที่ 3.4 หน้าต่างแสดงผลหลังกดปุ่ม Queries

1.4) ปุ่ม Views แสดงเว็บไซต์ที่เคยเข้าดูทั้งหมดในโปรเจค

รูปที่ 3.5 หน้าต่างแสดงผลหลังกดปุ่ม Views

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

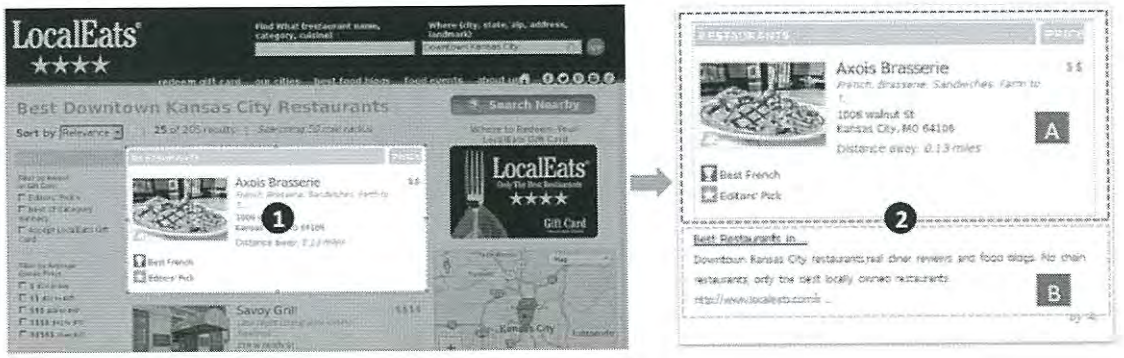
2. แถบสำหรับใส่คำที่ใช้ในการค้นหา และมีคำแนะนำสำหรับการค้นหา ซึ่งเราใช้ผลการค้นหาของ Bing
3. แถบสำหรับตัดสินว่าเว็บไซต์เกี่ยวข้องหรือไม่ โดยมี 2 ปุ่ม ได้แก่ ปุ่มระบุว่าเว็บไซต์เกี่ยวข้อง (thumbs-up) และปุ่มระบุว่าเว็บไซต์ไม่เกี่ยวข้อง (thumbs-down)
4. แถบแสดงตัวอย่างรูปภาพสรุปของเว็บไซต์ ซึ่งถูกตัดโดยผู้ค้นหา
5. แถบแสดงความคิดเห็น เพื่อให้ผู้ค้นหาแสดงความคิดเห็นเกี่ยวกับเว็บไซต์ (comment)
6. ขวาบนคือ กล่องแสดงกิจกรรมทั้งหมดที่เกิดขึ้นระหว่างการค้นหาข้อมูลแบบทันทีตลอดเวลา (real time)
7. ด้านล่าง-ขวา คือ กล่องสำหรับให้ผู้ค้นหาสนทนากันแบบทันทีตลอดเวลา (instant messaging)

3.3 กระดานแสดงและแบ่งปันรูปภาพภายในกลุ่ม (Snapboard)

รูปภาพขนาดย่อของเว็บไซต์ (thumbnail) ถูกนำมาใช้แทนข้อความตัวตัวอักษร (title, URL, textual snippet) เพื่อใช้เป็นตัวแทนเว็บไซต์มากขึ้นในงานด้าน "การปฏิสัมพันธ์ระหว่างมนุษย์และคอมพิวเตอร์" (Human Computer Interaction) และงานล่าสุดคือ "ภาพสรุปของเว็บไซต์" (visual snippet) ถูกพัฒนาขึ้นและเป็นที่น่าสนใจมากกว่าภาพขนาดย่อและข้อความอักษร ภาพสรุปบางส่วนของเว็บไซต์ เป็นการรวบรวมเฉพาะสิ่งที่สำคัญที่สื่อถึงเว็บไซต์นั้น มารวมไว้ในรูปภาพเดียวโดยระบบเอง ซึ่งอาจจะไม่ตรงกับข้อมูลที่ผู้ค้นหาต้องการจริง และในเว็บไซต์ที่มีข้อมูลตรงกับความต้องการ (information need) อาจจะตรงความต้องการแค่บางส่วนเท่านั้น ซึ่งไม่ตรงความต้องการทั้งเว็บไซต์ เราจึงนำเสนอสรุปแบบภาพ ซึ่งเกิดจากระบบที่เราพัฒนาขึ้น มีฟังก์ชันให้ผู้ค้นหาสามารถตัดข้อมูลบางส่วนในเว็บไซต์ที่ตรงกับความต้องการด้วยตัวเอง นอกจากนี้อย่างที่ทราบกันแล้วว่า รูปภาพขนาดย่อของเว็บไซต์ทำให้ผู้ค้นหาประเมินความเกี่ยวข้องของเว็บไซต์นั้นต่ำเกินไป และข้อความสรุปแบบตัวอักษรทำให้ผู้ค้นหาประเมินความเกี่ยวข้องของเว็บไซต์นั้นสูงเกินไป ทำให้เรากังวลว่าจะเจอปัญหาเดียวกัน ดังนั้น สรุปแบบภาพของเรา จึงประกอบด้วยรูปภาพที่ผู้ค้นหาตัดและข้อความสรุปแบบตัวอักษรด้วย

รูปที่ 3.5 แสดงตัวอย่างการตัดภาพ (1) และองค์ประกอบของภาพหลังจากที่ผู้ใช้ตัดแล้ว (2) ซึ่งเป็นการรวมกันของรูปภาพที่ตัดกับข้อมูลตัวอักษร (title, textual snippet, URL) ของเว็บไซต์นั้น เมื่อผู้ค้นหาคลิกรูปภาพที่ (2A) รูปภาพจะขยายใหญ่ขึ้นให้เห็นภาพชัดเจน เมื่อผู้ใช้คลิกข้อมูลที่ (2B) จะเปิดหน้าเว็บไซต์ของรูปที่ผู้ค้นหาตัดในระบบโคสเปช

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 ตัวอย่าง (1) การตัดสรุปแบบภาพ โดยผู้ใช้ (2) การทำสรุปแบบภาพโดยการรวมกันของภาพที่ผู้ใช้ตัดและข้อความสรุปเว็บไซต์และตัวอักษร

ในระบบโคสเปช ภาพสรุปของเว็บไซต์ ถูกแสดงในกระดานที่เรียกว่า กระดานแสดงและแบ่งปันรูปภาพเฉพาะกลุ่ม (Snapboard) ซึ่งมี 2 ชนิดคือ 1) กระดานแสดงรูปภาพสรุปของเว็บไซต์ทั้งหมดในโปรเจค ซึ่งแสดงในรูปที่ 3.6 และ 2) กระดานแสดงรูปภาพสรุปแยกของแต่ละเว็บไซต์ ซึ่งแสดงในรูปที่ 3.1 (4)

รูปที่ 3.7 ตัวอย่างฟังก์ชัน สแนปบอร์ด กระดานแสดงและแบ่งปันรูปภาพภายในกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การศึกษาระบบโคสเปซจากกลุ่มผู้ใช้งานในเรื่องการค้นหาแบบคนเดียวและค้นหาร่วมกันแบบต่างเวลา

ในบทนี้เรานำเสนอการศึกษากลุ่มผู้ใช้งานระบบค้นหาข้อมูลแบบร่วมมือกันชื่อ "โคสเปซ" (CoZpace) เป็นการประเมินประสิทธิภาพของเครื่องมือที่ช่วยในการค้นหาร่วมกันภายในระบบ โดยเปรียบเทียบระหว่าง 2 ระบบ คือ 1) ระบบที่มีฟังก์ชันการตัดภาพสรุปของเว็บไซต์ 2) ระบบที่มีฟังก์ชันแสดงความคิดเห็นของเว็บไซต์

4.1 ปัญหางานวิจัย

ปัจจุบันรูปภาพถูกนำมาช่วยในการสื่อสารระหว่างค้นหาข้อมูลแบบร่วมมือมากขึ้น แต่ยังรองรับการค้นหาข้อมูลพร้อมในเวลาเดียวกันเท่านั้น เราจึงนำเสนอฟังก์ชันใหม่ชื่อ "กระดานแสดงและแบ่งปันภาพเฉพาะกลุ่ม" (Snapboard) ซึ่งช่วยผู้ใช้ในการค้นหาข้อมูลแบบร่วมมือแบบต่างเวลา เพื่อประเมินประสิทธิภาพของฟังก์ชันใหม่ เนื่องจากการแสดงความคิดเห็นและการบันทึกเว็บไซต์ที่เกี่ยวข้องเป็นฟังก์ชันที่ผู้ค้นหานิยมทำมากที่สุดในการค้นหาข้อมูลร่วมกันแบบต่างเวลา เราจัดทำการทดลองตามหลักการศึกษาระบบค้นหาข้อมูลแบบมีปฏิสัมพันธ์กับกลุ่มผู้ใช้งาน ที่แนะนำโดย P. Borlund [11] และมีการเก็บข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพจากกิจกรรมที่กลุ่มผู้ทดลองได้ทำระหว่างค้นหา โดยเราจำลองสถานการณ์ในการค้นหาข้อมูลออกเป็น 2 สถานการณ์ คือ 1) เมื่อผู้ทดลองต้องค้นหาข้อมูลเพียงคนเดียว และ 2) เมื่อผู้ทดลองต้องค้นหาข้อมูลร่วมกันแบบต่างเวลากัน

4.2 ขอบเขตของการศึกษา

1. ผลการค้นหาเป็นข้อมูลประเภทเว็บไซต์เท่านั้น
2. ศึกษาในเรื่องส่วนต่อประสานกับผู้ใช้ (user interface)
3. ศึกษาระบบโคสเปซ ในเรื่องการร่วมมือกันในการค้นหาข้อมูลแบบตั้งใจของผู้ใช้
4. ใช้ผลการค้นหาของ Microsoft Bing ในการศึกษาฟังก์ชันสแนปบอร์ดจากกลุ่มผู้ใช้
5. ศึกษากลุ่มผู้ใช้ทั้งในการค้นหาข้อมูลแบบคนเดียวและแบบร่วมมือกันในเวลาที่ต่างกัน
6. เปรียบเทียบฟังก์ชันการตัดภาพสรุปของเว็บไซต์กับฟังก์ชันแสดงความคิดเห็นของเว็บไซต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. เก็บข้อมูลทั้งในรูปเชิงปริมาณและเชิงคุณภาพเพื่อวิเคราะห์ผลการทดลอง

4.3 สมมติฐานงานวิจัย

1. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ค้นหาในการค้นหางานที่คนมักค้นหาร่วมกัน
2. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ใช้ในการค้นหาคนเดียว
3. กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ใช้ในการค้นหาร่วมกันแบบต่างเวลา

4.4 ผู้เข้าร่วมการทดลอง

ผู้เข้าร่วมการทดลองเป็นอาสาสมัคร จำนวน 32 คน โดยเป็นนักเรียนมัธยมศึกษาปีที่ 5 ภาคภาษาอังกฤษ จากโรงเรียนนครนายกวิทยาคม จำนวน 12 คน และเป็นนักศึกษาระดับปริญญาตรี จำนวน 10 คน ระดับปริญญาโท 12 คน จากคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง พวกเขาทั้งหมดมีทักษะด้านภาษาอังกฤษและการค้นหาเว็บไซต์ จากแบบสอบถามก่อนการทดลองเราพบว่า พวกเขาส่วนใหญ่เคยค้นหาข้อมูลบนเว็บไซต์ร่วมกับเพื่อน เฉลี่ยกลุ่มละประมาณ 3 คน โดยใช้เครือข่ายสังคมหรืออีเมล์ในการติดต่อสื่อสาร แต่พวกเขาไม่เคยใช้ระบบค้นหาข้อมูลแบบร่วมมือกันมาก่อน พวกเขาทั้งหมดถูกจับคู่เป็น 16 คู่ ใน เพื่อทำการทดลองการค้นหาข้อมูลในเวลาที่ต่างเวลากัน โดยเรากำหนดผู้ค้นหาแรกเป็น P1 คือ ผู้เริ่มค้นหาที่ไม่มีข้อมูลช่วยในการค้นหาเลย และ P2 คือผู้ค้นหาที่ทำการค้นหาต่อจาก P1 ในกลุ่มนี้จะมีข้อมูลของ P1 ช่วยในการค้นหา โดยในการทดลองมีข้อจำกัดเรื่องความรู้ของกลุ่มนักเรียนและนักศึกษาที่มีความแตกต่างกัน ซึ่งอาจส่งผลกระทบต่อผลการทดลอง

รูปที่ 4.1 ผู้เข้าร่วมการทดลองเป็นนักเรียนจากโรงเรียนนครนายกวิทยาคม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 โจทย์ในการทำการทดลอง

โจทย์ในการทำการทดลอง เราได้เลือกประเภทของโจทย์ ที่คนมักค้นหาร่วมกันมากที่สุด 4 อันดับแรกมาใช้ในการทดลอง ได้แก่ การวางแผนเดินทาง การหาข้อมูลซื้อของทั่วไป การค้นหางานวรรณกรรม และการค้นหาข้อมูลด้านเทคนิค ซึ่งนำเสนอโดย Morris [1]

4.5.1 การวางแผนเดินทาง (Travel Planning)

T1: ให้ผู้เข้าร่วมการทดลองจินตนาการว่า พวกเขาวางแผนเดินทางไปเมืองแคนซัส ประเทศสหรัฐอเมริกา พวกเขาต้องค้นหาข้อมูลเกี่ยวกับการไปเที่ยวช่วงวันหยุดของพวกเขา ตัวอย่างเป้าหมายที่ต้องค้นหา เช่น ไปเมืองแคนซัสอย่างไร พักที่ไหน หรือทำอะไรบ้าง เป็นต้น

4.5.2 การหาข้อมูลซื้อของทั่วไป (General Shopping)

T2: ให้ผู้เข้าร่วมการทดลองจินตนาการว่า พวกเขาได้รับเงินจำนวน 30,000 ดอลลาร์สหรัฐ เพื่อซื้อรถยนต์จำนวน 1 คัน เป้าหมายของพวกเขา เช่น การหาข้อมูลทางเทคนิคของรถยนต์ ยี่ห้อของรถยนต์ หรือร้านค้าที่ขายรถยนต์ เป็นต้น

4.5.3 การค้นหางานวรรณกรรม (Literature Search)

T3: ให้ผู้เข้าร่วมการทดลองจินตนาการว่า พวกเขาได้รับมอบหมายให้เขียนบทความเกี่ยวกับสงครามกลางเมืองสหรัฐ งานของพวกเขาคือการหาข้อมูล เช่น สาเหตุของการเกิดสงครามกลางเมือง สาเหตุทางเศรษฐกิจ ผลกระทบของสงครามกลางเมืองที่ส่งผลในปัจจุบัน หรืออาวุธที่ใช้ในระหว่างสงครามกลางเมือง เป็นต้น

4.5.4 การค้นหาข้อมูลด้านเทคนิค (Technical Information)

T4: ให้ผู้เข้าร่วมการทดลองจินตนาการว่า พวกเขาต้องการที่จะลดการใช้เครื่องปรับอากาศภายในบ้าน งานของพวกเขาคือการค้นหาวัสดุที่ช่วยเคลือบหลังคาบ้าน หรือวิธีการแก้ปัญหาเกี่ยวกับการทำให้อุณหภูมิภายในบ้านลดลง เป็นต้น

4.6 สถานการณ์จำลองในการทดลอง

เราศึกษาประสิทธิผลของฟังก์ชันตัดภาพสรุปหน้าเว็บไซต์และกระดานแสดงและแบ่งปันภาพภายในกลุ่ม โดยการเปรียบเทียบระหว่าง 2 ระบบ คือ S1) ระบบที่มีฟังก์ชันการตัดภาพสรุปของเว็บไซต์ และ S2) ระบบที่มีฟังก์ชันแสดงความคิดเห็นของเว็บไซต์ ในสองสถานการณ์จำลองสำหรับการค้นหาข้อมูลดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.1 การค้นหาเพียงคนเดียว (Individual Search)

IN: ในสถานการณ์นี้ กลุ่มผู้ทดลอง P1 จะได้รับมอบหมายให้ค้นหาข้อมูลเพียงคนเดียว พวกเขาต้องหาเว็บไซต์ที่เกี่ยวข้องกับโจทย์ให้ได้มากที่สุด และต้องระบุว่าทำไมเว็บไซต์นั้นถึงเกี่ยวข้อง โดยการตัดสรุปแบบภาพ ในระบบ S1 หรือการแสดงความคิดเห็นในระบบ S2

4.6.2 การค้นหาร่วมกันในเวลาที่ต่างกัน (Asynchronous Collaborative Search)

AC: ในสถานการณ์นี้คล้าย IN คือกลุ่มผู้ทดลอง P1 จะได้รับมอบหมายให้ค้นหาข้อมูลก่อน พวกเขาต้องหาเว็บไซต์ที่เกี่ยวข้องกับโจทย์ให้ได้มากที่สุด และต้องระบุว่าทำไมเว็บไซต์นั้นถึงเกี่ยวข้อง โดยการตัดสรุปแบบภาพ ในระบบ S1 หรือการแสดงความคิดเห็นในระบบ S2 หลังจาก P1 ค้นหาเสร็จแล้ว กลุ่มผู้ทดลอง P2 จะได้รับมอบหมายให้ค้นหาข้อมูลต่อจาก P1 และทำเช่นเดียวกัน แตกต่างกันที่ P2 มีข้อมูลของ P1 ช่วยในการค้นหา หลังจากที่ P2 ค้นหาเสร็จแล้ว ทั้ง P1 และ P2 จะได้รับมอบหมายให้ตัดสินเว็บไซต์ที่เพื่อนตัดสินว่าเกี่ยวข้องแต่ตัวเองยังไม่ได้ตัดสินว่าเกี่ยวข้องหรือไม่ โดยให้ตัดสินจากเหตุผลที่เพื่อนระบุโดยสรุปแบบภาพ ในระบบ S1 หรือความคิดเห็นในระบบ S2 ซึ่งเราเรียกว่า "Review session" ซึ่งทำให้เรามั่นใจว่าผู้ทดลองทั้ง P1 และ P2 จะได้รับประสบการณ์ในการใช้ทั้งฟังก์ชันกระดานแสดงและแบ่งปันสรุปแบบภาพ ภายในกลุ่มและฟังก์ชันแสดงความคิดเห็น เพื่อทำแบบสอบถามเชิงคุณภาพ

4.7 การเก็บข้อมูลสำหรับการศึกษา

เพื่อตอบคำถามงานวิจัยเราได้ทำการเก็บข้อมูลทั้งในเชิงปริมาณ (quantitative data) และเชิงคุณภาพ (qualitative data)) ตารางที่ 4.1 แสดงข้อมูลเชิงปริมาณ ซึ่งเราเก็บบันทึกจากจำนวนกิจกรรมระหว่างการค้นหาของผู้ทดลอง (logs) ส่วนการเก็บข้อมูลเชิงคุณภาพ เราใช้ 2 วิธีคือ 1) แบบสอบถามคำถามปลายเปิด-ปิด (open-ended questionnaires) โดยให้ผู้ทดลองให้คะแนนแบบ 5-point Likert scale และ 2) การสนทนากลุ่ม (Focus Group Discussion) เป็นการรวบรวมข้อมูลจากการสนทนากับกลุ่มผู้ทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 แสดงข้อมูลเชิงปริมาณที่เก็บบันทึกเพื่อการวิเคราะห์ผลการทดลอง

Actions	Descriptions
Query (Query)	จำนวนคำค้นหา ที่ผู้ทดลองใช้ในการค้นหา
View (Viewed Web)	จำนวนเว็บไซต์ทั้งหมด ที่ผู้ทดลองเข้าดู
Relevant (Rel Web)	จำนวนเว็บไซต์ที่เกี่ยวข้อง ซึ่งระบุโดยผู้ทดลอง
Visual snippet (Vsnip)	จำนวนสรุปแบบภาพ ซึ่งถูกตัดโดยผู้ทดลอง
Comment (Comt)	จำนวนความคิดเห็น ที่ผู้ทดลองระบุในเว็บไซต์

4.8 ขั้นตอนการทดลอง

ในการทดลอง แต่ละคู่จะได้ทำการค้นหาข้อมูลแบบร่วมมือกันในเวลาที่ต่างกัน ในโจทย์ทั้ง 4 ข้อ โดยใช้ 2 ระบบ คือ S1) ระบบแบบมีฟังก์ชันตัดสรุปแบบภาพ และ S2) ระบบแบบไม่มีฟังก์ชันตัดสรุปแบบภาพ เราออกแบบการทดลองตามหลัก Graeco-Latin Square ซึ่งแสดงในตารางที่ 4.2 โดยเป็นการหมุนเวียนลำดับการทำโจทย์และระบบที่ใช้ เพื่อควบคุมผลกระทบที่เกี่ยวกับการเรียนรู้ที่เพิ่มขึ้นของผู้ทดลอง ซึ่งอาจทำให้ผู้ทดลองทำโจทย์ข้อหลังหรือระบบหลังได้ดีกว่าโจทย์ข้อแรกหรือระบบแรกเสมอ เนื่องจากธรรมชาติในการเรียนรู้ที่เพิ่มขึ้นของมนุษย์

การทดลองมีระยะเวลาประมาณ 3 ชั่วโมง เริ่มต้นเราทำการทดลองแบบค้นหาคนเดียว โดยเราจะอธิบายรายละเอียดในการทำการทดลอง และให้ผู้ทดลองเซ็นสัญญายินยอมเข้าร่วมการทดลองก่อนจะเริ่มทำแบบสอบถามข้อมูลส่วนตัวก่อนทำการทดลอง ต่อจากนั้นผู้ทดลองจะได้รับการสอนการใช้งานระบบและทดลองใช้ระบบประมาณ 15 นาที เพื่อให้เกิดความคุ้นเคย หลังจากทดลองใช้ระบบเสร็จ ผู้ทดลองจะได้ทำการค้นหาตามโจทย์ที่ได้รับทั้ง 4 ข้อ โดยโจทย์แต่ละข้อจะให้เวลาในการทำ 15 นาที หลังจากค้นหาโจทย์แต่ละเสร็จ ผู้ทดลองจะได้รับแบบสอบถามหลังทำโจทย์ ซึ่งถามเกี่ยวกับประสบการณ์ในการค้นหาข้อมูลในโจทย์นั้น หลังจากทำการค้นหาเสร็จทั้ง 4 โจทย์ ผู้ทดลองจะได้รับแบบสอบถามหลังการทดลองเพื่อตัดสินว่า ผู้ทดลองพึงพอใจกับระบบแบบไหนมากกว่ากัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงตัวอย่างการหมุนเวียนลำดับการใช้ระบบ S1 และ S2 กับการทำโจทย์ T1, T2, T3 และ T4 ของผู้ทดลองทั้งหมด 32 คน

	S1	S2	**S1**	S2
P1, P2	**T1**	T2	T3	T4
P3, P4	T2	T3	T4	**T1**
P5, P6	T3	T4	**T1**	T2
P7, P8	T4	**T1**	T2	T3

	S1	S2	**S1**	S2
P9, P10	T1	**T2**	T3	T4
P11, P12	**T2**	T3	T4	T1
P13, P14	T3	T4	T1	**T2**
P15, P16	T4	T1	**T2**	T3

	S2	S1	**S2**	S1
P17, P18	T1	T2	**T3**	T4
P19, P20	T2	**T3**	T4	T1
P21, P22	**T3**	T4	T1	T2
P23, P24	T4	T1	T2	**T3**

	S2	S1	**S2**	S1
P25, P26	T1	T2	T3	**T4**
P27, P28	T2	T3	**T4**	T1
P29, P30	T3	**T4**	T1	T2
P31, P32	**T4**	T1	T2	T3

4.9 ผลการทดลอง

4.9.1 ประสิทธิภาพของการตัดภาพสรุปของเว็บไซต์ในการค้นหาคนเดียวและค้นหาร่วมกัน

ตารางที่ 4.3 แสดงการเปรียบเทียบประสิทธิภาพระหว่าง 2 ระบบ คือ ระบบที่มีฟังก์ชันการตัดภาพสรุปของเว็บไซต์ (S1) และระบบที่มีฟังก์ชันแสดงความคิดเห็นของเว็บไซต์ (S2) ใน 2 เงื่อนไข คือ การค้นหาข้อมูลคนเดียว (IN) และการค้นหาข้อมูลร่วมกันแบบต่างเวลากัน (AC) เราพบว่าจำนวนเฉลี่ยการตัดภาพสรุปของเว็บไซต์ (Vsnip) ในระบบ S1 มีมากกว่าจำนวนเฉลี่ยการแสดงความคิดเห็น (Comt) ในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งหมายความว่า ผู้ทดลองชอบการสรุปเว็บไซต์แบบการตัดภาพมากกว่าการแสดงความคิดเห็นของเว็บไซต์ และจำนวนเฉลี่ยคำที่ใช้ค้นหาในระบบ S1 มากกว่าระบบ S2 ทั้งในเงื่อนไข IN และ AC แต่ไม่สามารถสรุปได้ว่าผู้ทดลองต้องพยายามคิดคำที่ใช้สำหรับค้นหาในระบบ S1 มากกว่าระบบ S2 เพราะผู้ทดลองได้แสดงความเห็นว่า โจทย์ที่ใช้ในการทดลองง่ายต่อการสร้างคำที่ใช้สำหรับค้นหา ซึ่งแสดงในตารางที่ 4.6

นอกจากนี้เรายังพบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในการค้นคืนเว็บไซต์ที่เกี่ยวข้อง (All Rel) ซึ่งใน AC จำนวนเว็บไซต์ที่เกี่ยวข้อง นับเฉพาะจากเว็บไซต์ที่ทั้งผู้ทดลอง P1 และ P2 ระบุว่าเกี่ยวข้อง ในกลุ่มเดียวกันเพื่อให้มั่นใจว่าเป็นเว็บไซต์ที่เกี่ยวข้องจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรายังมีการคำนวณจำนวนเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) และพบว่ามีจำนวนแตกต่างจากจำนวนเว็บไซต์ที่เกี่ยวข้องทั้งหมด (All Rel) ทั้งใน IN และ AC ทำให้เรามั่นใจว่าแต่ละกลุ่มมีการเลือกเว็บไซต์เดียวกันเกิดขึ้น แสดงว่าเว็บไซต์นั้นน่าจะมีความเกี่ยวข้องจริง

เราตรวจสอบเปอร์เซ็นต์ของเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) ต่อเว็บไซต์ที่ผู้ค้นหาดูและไม่ซ้ำ (Uniq View) พบว่าในระบบ S1 สูงกว่าในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งทำให้เราตอบสมมติ ข้อที่ 2) และ 3) ได้ว่า กระดานแสดงและแบ่งปันภาพสรุป ช่วยผู้ค้นหาทั้งในการค้นหาคนเดียวและในการค้นหาร่วมกันแบบต่างเวลากัน

เรามีการวิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ โดยการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ในเงื่อนไข IN ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel และ All View ที่ $p < 0.1$ พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม Vsnip with Comt โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.035*$ และในกิจกรรม Query, All Rel, All View มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.188$, $p = 0.634$, และ $p = 0.966$ ตามลำดับ

การวิเคราะห์ความแปรปรวนแบบทางเดียวยังถูกนำมาใช้วิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ ในเงื่อนไข AC ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel, และ All View ที่ $p < 0.1$ ด้วย พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม Vsnip with Comt และ All Rel โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.095*$ และ $p = 0.004*$ ตามลำดับ และในกิจกรรม Query และ All View มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.395$ และ $p = 0.337$ ตามลำดับ

ตารางที่ 4.3 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่ผู้ทดลองกระทำในการค้นหาข้อมูล (เฉลี่ย 16 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันแบบต่างเวลา (AC) ในส่วนของวงเล็บคือค่า S.D.

					Rel Web		Viewed Web		% Uniq Rel
Cond	Sys	Vsnip	Comt	Query	All Rel	Uniq Rel	All View	Uniq View	per Uniq View
IN	S1	**20.81*** (4.25)	-	**13.69** (3.57)	**12.50** (3.33)	**8.44**	**23.19** (5.15)	14.38	**58.70**
	S2	-	15.88 (5.23)	11.31 (2.72)	11.75 (4.14)	7.88	23.06 (7.99)	**14.44**	54.54
AC	S1	**33.50*** (4.37)	-	**22.38** (3.80)	**25.75*** (3.61)	**13.88**	**53.50** (6.18)	**26.69**	**52.26**
	S2	-	28.25 (4.64)	20.44 (2.78)	18.69 (3.46)	10.44	49.31 (7.42)	25.31	41.25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้เรายังมีการวิเคราะห์ผลการทดลองแยกตามประเภทผู้ค้นหาคือ 1) นักเรียน และ 2) นักศึกษา ตารางที่ 4.4 เปรียบเทียบ 2 ระบบในกลุ่มนักเรียนพบว่า จำนวนเฉลี่ยการตัดภาพสรุปของเว็บไซต์ (Vsnip) ในระบบ S1 มีมากกว่าจำนวนเฉลี่ยการแสดงความคิดเห็น (Comt) ในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งหมายความว่า กลุ่มนักเรียนชอบการสรุปเว็บไซต์แบบการตัดภาพมากกว่าการแสดงความคิดเห็นของเว็บไซต์ และจำนวนเฉลี่ยคำที่ใช้ค้นหาในระบบ S1 มากกว่าระบบ S2 ทั้งในเงื่อนไข IN และ AC แต่ไม่สามารถสรุปได้ว่ากลุ่มนักเรียนต้องพยายามคิดคำที่ใช้ค้นหาในระบบ S1 มากกว่าระบบ S2 เพราะผู้ทดลองได้แสดงความเห็นว่า โจทย์ที่ใช้ในการทดลองง่ายในการสร้างคำที่ใช้สำหรับค้นหา ซึ่งแสดงในตารางที่ 4.6

นอกจากนี้เรายังพบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในการค้นคืนเว็บไซต์ที่เกี่ยวข้อง (All Rel) ซึ่งใน AC จำนวนเว็บไซต์ที่เกี่ยวข้อง นับเฉพาะจากเว็บไซต์ที่ทั้งผู้ทดลอง P1 และ P2 ระบุว่าเกี่ยวข้อง ในกลุ่มเดียวกันเพื่อให้มั่นใจว่าเป็นเว็บไซต์ที่เกี่ยวข้องจริง

เรายังมีการคำนวณจำนวนเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) และพบว่ามีจำนวนแตกต่างจากจำนวนเว็บไซต์ที่เกี่ยวข้องทั้งหมด (All Rel) ทั้งใน IN และ AC ทำให้เรามั่นใจว่าแต่ละกลุ่มมีการเลือกเว็บไซต์เดียวกันเกิดขึ้น แสดงว่าเว็บไซต์นั้นน่าจะมีความเกี่ยวข้องจริง

เราตรวจสอบเปอร์เซ็นต์ของเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) ต่อเว็บไซต์ที่ผู้ค้นหาดูและไม่ซ้ำ (Uniq View) พบว่าในระบบ S1 สูงกว่าในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งทำให้เราตอบสมมติ ข้อที่ 2) และ 3) ได้ว่า กระดานแสดงและแบ่งปันภาพสรุป ช่วยผู้ค้นหาทั้งในการค้นหาคนเดียวและในการค้นหาร่วมกันแบบต่างเวลากัน

นอกจากนี้เรามีการวิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติในกลุ่มนักเรียน โดยการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ในเงื่อนไข IN ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel และ All View ที่ $p < 0.1$ พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม Vsnip with Comt โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.010*$ และในกิจกรรม Query, All Rel, All View มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.207$, $p = 0.254$, และ $p = 0.210$ ตามลำดับ

การวิเคราะห์ความแปรปรวนแบบทางเดียวยังถูกนำมาใช้วิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ ในเงื่อนไข AC ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel, และ All View ที่ $p < 0.1$ ด้วย พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม Vsnip with Comt, All Rel, และ All View โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.006*$, $p = 0.016*$, และ $0.094*$ ตามลำดับ และในกิจกรรม Query มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.328$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่นักเรียนกระทำในการค้นหาข้อมูล (เฉลี่ย 6 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันแบบต่างเวลา (AC) ในส่วนของวงเล็บคือค่า S.D.

					Rel Web		Viewed Web		% Uniq Rel
Cond	Sys	Vsnip	Comt	Query	All Rel	Uniq Rel	All View	Uniq View	per Uniq View
IN	S1	**17.50*** (4.09)	-	**13.50** (4.25)	**12.67** (3.85)	**5.83**	**23.00** (5.37)	10.50	**55.52**
	S2	-	12.16 (3.32)	9.50 (2.41)	10.00 (3.22)	4.50	18.00 (5.58)	**11.83**	38.04
AC	S1	**30.67*** (3.64)	-	**21.67** (4.18)	**24.83*** (3.86)	**11.33**	**51.50*** (5.83)	20.22	**56.03**
	S2	-	21.50 (2.87)	17.67 (2.54)	15.33 (2.63)	6.66	40.83 (5.16)	**28.83**	23.10

ตารางที่ 4.5 เปรียบเทียบ 2 ระบบในกลุ่มนักศึกษาพบว่า จำนวนเฉลี่ยการตัดภาพสรุปของเว็บไซต์ (Vsnip) ในระบบ S1 มีมากกว่าจำนวนเฉลี่ยการแสดงความคิดเห็น (Comt) ในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งหมายความว่า กลุ่มนักศึกษาชอบการสรุปเว็บไซต์แบบการตัดภาพมากกว่าการแสดงความคิดเห็นของเว็บไซต์ และจำนวนเฉลี่ยคำที่ใช้ค้นหาในระบบ S1 มากกว่าระบบ S2 ทั้งในเงื่อนไข IN และ AC แต่ไม่สามารถสรุปได้ว่ากลุ่มนักศึกษาต้องพยายามคิดคำที่ใช้ในค้นหาในระบบ S1 มากกว่าระบบ S2 เพราะผู้ทดลองได้แสดงความเห็นว่า โจทย์ที่ใช้ในการทดลองง่ายในการสร้างคำที่ใช้สำหรับค้นหา ซึ่งแสดงในตารางที่ 4.6

นอกจากนี้เรายังพบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในการค้นคืนเว็บไซต์ที่เกี่ยวข้อง (All Rel) ซึ่งใน AC จำนวนเว็บไซต์ที่เกี่ยวข้อง นับเฉพาะจากเว็บไซต์ที่ทั้งผู้ทดลอง P1 และ P2 ระบุว่าเกี่ยวข้อง ในกลุ่มเดียวกันเพื่อให้มั่นใจว่าเป็นเว็บไซต์ที่เกี่ยวข้องจริง

เรายังมีการคำนวณจำนวนเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) และพบว่ามีจำนวนแตกต่างจากจำนวนเว็บไซต์ที่เกี่ยวข้องทั้งหมด (All Rel) ทั้งใน IN และ AC ทำให้เรามั่นใจว่าแต่ละกลุ่มมีการเลือกเว็บไซต์เดียวกันเกิดขึ้น แสดงว่าเว็บไซต์นั้นน่าจะมีความเกี่ยวข้องจริง

เราตรวจสอบเปอร์เซ็นต์ของเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) ต่อเว็บไซต์ที่ผู้ค้นหาดูและไม่ซ้ำ (Uniq View) พบว่าในระบบ S1 สูงกว่าในระบบ S2 ทั้งในเงื่อนไข IN และ AC ซึ่งทำให้เราตอบสมมติ ข้อที่ 2) และ 3) ได้ว่า กระดานแสดงและแบ่งปันภาพสรุป ช่วยผู้ค้นหาทั้งในการค้นหาคนเดียวและในการค้นหาร่วมกันแบบต่างเวลากัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้เรามีการวิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติในกลุ่มนักศึกษา โดยการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) ในเงื่อนไข IN ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel และ All View ที่ $p < 0.1$ พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel, All View มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.151$, $p = 0.328$, $p = 0.853$, และ $p = 0.483$ ตามลำดับ

การวิเคราะห์ความแปรปรวนแบบทางเดียวยังถูกนำมาใช้วิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ ในเงื่อนไข AC ของระบบ S1 และระบบ S2 ในกิจกรรม Vsnip with Comt, Query, All Rel, และ All View ที่ $p < 0.1$ ด้วย พบว่าระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในกิจกรรม All Rel โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.074$* และในกิจกรรม Vsnip with Comt, Query, และ All View มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.512$, $p = 0.798$, และ $p = 0.958$ ตามลำดับ

ตารางที่ 4.5 แสดงค่าเฉลี่ยจำนวนกิจกรรมที่นักศึกษากระทำในการค้นหาข้อมูล (เฉลี่ย 10 คน/กลุ่ม) ในเหตุการณ์การค้นหาเพียงคนเดียว (IN) และการค้นหาร่วมกันแบบต่างเวลา (AC) ในส่วนของวงเล็บคือค่า S.D.

					Rel Web		Viewed Web		% Uniq Rel
Cond	Sys	Vsnip	Comt	Query	All Rel	Uniq Rel	All View	Uniq View	per Uniq View
IN	S1	**22.80** (4.12)	-	**13.80** (3.21)	12.40 (3.09)	**5.50**	23.30 (5.15)	**14.30**	**38.46**
	S2	-	18.10 (5.89)	13.10 (2.72)	**12.80** (4.60)	4.30	**26.10** (8.92)	11.80	36.44
AC	S1	**35.20** (4.75)	-	**22.80** (3.59)	**26.30*** (3.49)	**9.20**	**54.70** (6.43)	**23.20**	**39.66**
	S2	-	32.30 (5.21)	22.10 (2.87)	20.70 (3.82)	7.80	54.40 (8.30)	22.60	34.51

จากการแยกวิเคราะห์ผลการทดลองตามประเภทกลุ่มผู้ใช้งานทั้ง 2 ระบบ พบว่าการตัดสรุปแบบภาพ ช่วยกลุ่มผู้ค้นหาทั้งประเภทนักเรียนและนักศึกษา โดยเราพบปริมาณการตัดรูปภาพและการแสดงความคิดเห็นในกลุ่มนักศึกษาไม่ต่างกันมากและไม่มีนัยสำคัญทางสถิติ แต่ในกลุ่มนักเรียนต่างกันอย่างมากและมีนัยสำคัญทางสถิติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้เรายังพบว่า เปอร์เซ็นต์ของเว็บไซต์ที่เกี่ยวข้องและไม่ซ้ำ (Uniq Rel) ต่อเว็บไซต์ที่ผู้ค้นหาเข้าดูและไม่ซ้ำ (Uniq View) ของระบบ S1 และ S2 ทั้งใน IN และ AC ในกลุ่มนักเรียนแตกต่างกันอย่างมาก แต่ในกลุ่มนักศึกษาแตกต่างกันเล็กน้อยเท่านั้น เราจึงคิดว่าฟังก์ชันการตัดสรุปแบบภาพ มีประโยชน์อย่างมากในการค้นหาแบบร่วมมือในกลุ่มนักเรียน แต่ในกลุ่มนักศึกษาเราคิดว่าทั้งการตัดสรุปแบบภาพและการแสดงความคิดเห็นมีประโยชน์ทั้งคู่ แต่เนื่องจากข้อจำกัดเกี่ยวกับโจทย์ที่ใช้ค้นหา ซึ่งอาจไม่น่าสนใจต่อกลุ่มนักเรียน จึงอาจส่งผลกระทบทำให้กลุ่มนักเรียนเลือกที่จะใช้ฟังก์ชันการตัดรูปภาพมากกว่าการแสดงความคิดเห็น เมื่อกลุ่มนักเรียนไม่ได้มีความตั้งใจที่จะบรรลุเป้าหมายในการค้นหาจริง ๆ จึงอาจทำให้พวกเขาเลือกที่จะใช้ฟังก์ชันที่สะดวกในการค้นหามากกว่า โดยไม่คำนึงถึงประโยชน์ของแต่ละฟังก์ชัน แต่ถ้าใช้โจทย์ที่มีความน่าสนใจต่อกลุ่มนักเรียนในการค้นหา เมื่อกลุ่มนักเรียนมีความต้องการที่จะค้นหาข้อมูลจริง ๆ พวกเขาอาจจะใช้ทั้งฟังก์ชันการตัดรูปภาพและแสดงความคิดเห็นมากพอ ๆ กัน เช่นเดียวกับกลุ่มนักศึกษา

4.9.2 ความพึงพอใจของผู้ใช้หลังการทำการทดลอง

ตารางที่ 4.6 แสดงค่าเฉลี่ยคะแนนความพึงพอใจของผู้ใช้หลังทำโจทย์แต่ละข้อเสร็จ จากการวิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ โดยการวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) ของระบบ S1 และระบบ S2 ในโจทย์ T1, T2, T3 และ T4 ที่ $p < 0.1$ พบว่า ผู้ทดลองพึงพอใจระบบ S1 มากกว่าระบบ S2 ในคำถาม Q1, Q2, Q3, Q4, และ Q5 โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.0568*$, $p = 0.0856*$, $p = 0.0829*$, $p = 0.0934*$ และ $p = 0.0053*$ ตามลำดับ

จากการสังเกตเราพบว่าโจทย์คือ ผลกระทบหลัก (Main Effect) ซึ่งมีผลกระทบต่อความพึงพอใจของผู้ใช้ดังนี้คือ ผู้ใช้จะพึงพอใจระบบ S1 มากกว่าระบบ S2 ในโจทย์ T1, T2 และ T4 แต่ผู้ใช้จะพึงพอใจระบบ S2 มากกว่าระบบ S1 ในโจทย์ T3 เราจึงลองวิเคราะห์ความแตกต่างที่มีนัยสำคัญทางสถิติ โดยการวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) ของระบบ S1 และระบบ S2 ในคำถาม Q1, Q2, Q3, Q4, และ Q5 ที่ $p < 0.1$ พบว่า ผู้ทดลองพึงพอใจระบบ S1 มากกว่าระบบ S2 ในโจทย์ T1 และ T2 โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.0042*$ และ $p = 0.0015*$ ตามลำดับ และในโจทย์ T4 มีค่าเฉลี่ยความแตกต่างอย่างไม่มีนัยสำคัญที่ $p = 0.1315$ และผู้ทดลองพึงพอใจระบบ S2 มากกว่าระบบ S1 ในโจทย์ T3 โดยมีค่าเฉลี่ยความแตกต่างอย่างมีนัยสำคัญที่ $p = 0.0634*$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลองพบว่า กระดานแสดงและแบ่งปันภาพสรุปภายในกลุ่ม ช่วยในการค้นหางานประเภท T1 และ T2 ซึ่งเราคิดว่าเป็นเพราะ การแบ่งปันข้อมูลรูปภาพสรุปช่วยในการค้นหาร่วมกันในงานประเภท การวางแผนท่องเที่ยวและการซื้อของทั่วไปที่มีราคาแพง เช่น การแบ่งปันสถานที่ท่องเที่ยว การแบ่งปันภาพสิ่งของที่แนะนำ เป็นต้น ทำให้กระดานแสดงและแบ่งปันภาพสรุปภายในกลุ่มมีประโยชน์ ซึ่งแตกต่างจากการค้นหางานด้านวรรณกรรม เราคิดว่าเพราะเป็นงานที่ผู้ค้นหาต้องการแบ่งปันข้อมูลที่เป็นข้อความตัวอักษรมากกว่ารูปภาพ ทำให้ระบบ S2 ได้คะแนนจากงานประเภทนี้มากกว่าระบบ S1 และสุดท้ายงานประเภท การค้นหาข้อมูลด้านเทคนิค พบว่าได้คะแนนความพึงพอใจในระบบ S1 สูงกว่าระบบ S2 แม้ไม่มีนัยสำคัญแต่ก็ถือเป็นเสียงตอบรับที่ดี จากคะแนนความพึงพอใจในตารางที่ 4.6 แสดงว่าผู้ใช้ชอบระบบ S1 มากกว่าระบบ S2 จึงตอบสมมติข้อที่ 1) ได้ว่า กระดานแสดงและแบ่งปันสรุปแบบภาพ ช่วยผู้ค้นหาในงานที่คนมักค้นหาร่วมกันบ่อย ๆ ในเรื่อง วางแผนท่องเที่ยว และการซื้อของ แต่ไม่ช่วยในการค้นหางานด้านวรรณกรรม

ตารางที่ 4.6 แสดงค่าเฉลี่ยคะแนนความพึงพอใจหลังทำโจทย์แต่ละข้อของผู้ทดลอง

Question	System	Mean				Mean	S.D.
		*T1	*T2	*T3	T4		
***Q1.** I feel that the search task	S1	**3.19**	**3.19**	2.13	**2.88**	**2.84**	1.12
was easy after finished it.	S2	2.63	2.31	**2.25**	2.69	2.47	1.42
***Q2.** I was relaxed while	S1	**3.94**	**3.81**	2.63	**3.52**	**3.52**	1.04
carrying out the search task.	S2	3.31	3.25	**3.31**	3.13	3.25	1.08
***Q3.** It was easy to formulate	S1	**4.06**	**4.06**	**4.19**	**4.00**	**4.08**	0.55
initial queries on these topics.	S2	3.63	3.88	3.94	3.88	3.83	0.72
***Q4.** I had enough time to do an	S1	**4.06**	**4.19**	3.19	**3.88**	**3.83**	1.10
effective search task.	S2	3.56	3.44	**3.63**	3.44	3.52	1.14
***Q5.** I have succeeded in my	S1	**4.13**	**4.00**	3.38	**4.38**	**3.97**	0.91
performance of the search task.	S2	3.13	3.38	**3.94**	3.75	3.55	0.87

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 แสดงคะแนนเฉลี่ยความพึงพอใจต่อส่วนติดต่อผู้ใช้ของระบบโคสเปช ซึ่งเก็บหลังทำการทดลองทั้ง 4 โจทย์เสร็จ ผู้ทดลองจะถูกถามเกี่ยวกับส่วนติดต่อผู้ใช้ (interface) ของระบบโคสเปช รวมทั้ง 3 เครื่องมือหลักที่ช่วยในการติดต่อสื่อสารได้แก่ การแบ่งปันเว็บไซต์ที่เกี่ยวข้อง การแบ่งปันความคิดเห็น และการแบ่งปันภาพสรุปของเว็บไซต์ พบว่า ผู้ทดลองพึงพอใจการแบ่งปันภาพสรุปของเว็บไซต์มากที่สุดที่เฉลี่ย 4.41 คะแนน แต่การแบ่งปันความคิดเห็นได้คะแนนเฉลี่ยน้อยสุดที่ 3.34 และส่วนติดต่อผู้ใช้ของระบบโคสเปช ให้ผลตอบรับไปในทางที่ดี โดยได้คะแนนเฉลี่ย 4.03 คะแนน

ตารางที่ 4.7 คะแนนเฉลี่ยความพึงพอใจที่มีต่อส่วนติดต่อผู้ใช้ของระบบโคสเปชหลังเสร็จสิ้นการทำการทดลองของผู้ทดลอง

Question	Mean	S.D.
1. The search interface was easy to use.	4.03	0.82
2. Relevant function was helpful in exploring web.	4.13	0.71
3. Comment function in S1 was helpful in exploring web.	3.34	1.24
4. Snapshot function in S2 was helpful in exploring web.	**4.41**	0.71

นอกจากนี้ เรามีการเปรียบเทียบมุมมองการรับรู้ของผู้ทดลองที่มีต่อทั้ง 2 ระบบหลังเสร็จสิ้นการทดลอง ซึ่งแสดงในตารางที่ 4.8 จากผู้เข้าร่วมการทดลองทั้งหมด 32 คน เราพบว่า โดยภาพรวมของระบบแล้วผู้ทดลองรู้สึกว่าระบบ S1 ดีกว่าระบบ S2 คิดเป็น 24 คน แต่ในเรื่องการเรียนรู้การใช้งานระบบ พบว่าไม่มีความแตกต่างกันของทั้งสองระบบ เมื่อถามผู้ทดลองในเรื่องของความง่ายในการค้นหาข้อมูล พบว่าผู้ทดลองส่วนใหญ่รู้สึกว่าระบบ S1 ช่วยให้ค้นหาข้อมูลได้ง่ายกว่า คิดเป็น 20 คน นอกจากนี้ยังพบว่า ผู้ทดลองชอบระบบ S1 มากกว่าระบบ S2 คิดเป็น 25 คน และหลังจากใช้ทั้งสองระบบแล้ว พบว่าผู้ทดลองรู้สึกว่าระบบ S1 เปลี่ยนมุมมองการรับรู้ที่มีต่อโจทย์ของพวกเขาให้กว้างขึ้น รูปภาพช่วยทำให้พวกเขาเข้าใจถึงเป้าหมายในการค้นหาข้อมูลมากขึ้นกว่าการอ่านเพียงข้อความตัวอักษร คิดเป็น 21 คน และสุดท้ายผู้ทดลองส่วนใหญ่เลือกระบบ S1 มีประสิทธิภาพดีกว่าระบบ S2 ในการค้นหาข้อมูล คิดเป็น 21 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 การเปรียบเทียบมุมมองการรับรู้ของผู้ใช้ที่มีต่อระบบหลังเสร็จสิ้นการทดลอง

Which systems...	**S1**	**S2**	=
did you find BEST overall?	**24**	4	4
did you find easier to LEARN TO USE?	9	6	**17**
did you find easier to USE?	**20**	7	5
did you PREFER?	**25**	3	4
did you find changed your perception of the task?	**21**	4	7
did you find more EFFECTIVE for the tasks you performed?	**21**	5	6
Percentage	**62.50%**	15.10%	22.40%

หลังเสร็จสิ้นการทดลอง เราจัดกลุ่มสนทนาขึ้นเพื่อให้ผู้ทดลองแลกเปลี่ยนความคิดเห็นเกี่ยวกับประโยชน์เครื่องมือการตัดและแบ่งปันภาพสรุปของเว็บไซต์ที่ได้ใช้ พบว่าผู้ทดลองส่วนใหญ่ให้ความเห็นว่า สรุปแบบภาพ นอกจากช่วยให้พวกเขาเข้าใจและเห็นภาพสิ่งที่เพื่อนต้องการจะสื่อให้พวกเขารู้ได้ง่ายขึ้นแล้ว ยังช่วยให้พวกเขาง่ายในการจะแบ่งปันสิ่งที่พวกเขาต้องการให้กับเพื่อนอีกด้วย เพียงแค่ตัดภาพบนเว็บไซต์แทนที่จะต้องพิมพ์บอกเพื่อนว่าตรงส่วนไหนของเว็บไซต์ที่ตรงกับความต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การสกัดคำจากรูปภาพของหน้าเว็บเพจ

ในบทนี้เราทำการศึกษาการสกัดคำจากภาพสรุปของเว็บไซต์เบื้องต้น เพื่อปูทางในการนำไปสร้างคำแนะนำสำหรับการค้นหา

5.1 ปัญหางานวิจัย

เพื่อที่จะใช้รูปภาพสรุปของเว็บไซต์ที่ผู้ใช้ตัดเองให้เกิดประโยชน์สูงสุด เราพบว่าในรูปภาพสรุปของเว็บไซต์ที่ผู้ใช้ตัดนั้นมีบางส่วนในภาพเป็นข้อความตัวอักษร ซึ่งข้อความเหล่านั้นน่าจะมีประโยชน์ในการนำมาใช้สร้างคำแนะนำสำหรับการค้นหา (query suggestion) แม้จะมีเครื่องมือในการเข้าถึงข้อมูลในเว็บไซต์ได้เช่น Lemur search toolkit [25] [26] แต่มันเป็นการดึงคำในเว็บไซต์ออกมาทั้งหมด อย่างไรก็ตามคำที่เราต้องการนั้นเป็นคำบางส่วนของเว็บไซต์เท่านั้น เราทำการศึกษาการสกัดคำจากภาพสรุปของเว็บไซต์เบื้องต้น เพื่อเป็นแนวทางในการนำไปสร้างคำแนะนำสำหรับการค้นหาเบื้องต้น โดยใช้ข้อมูลจาก Clueweb09 และเปรียบเทียบผลความน่าเชื่อถือในการสกัดคำจากรูปภาพกับ Lemur search toolkit โดยใช้ตัวชี้วัด Precision, Recall, และ F-measure

5.2 ขอบเขตของการศึกษา

1. ใช้เว็บเพจทั้งหมด 1,000 เว็บเพจ จาก Clueweb09 ในการทำการทดลองการสกัดคำจากสรุปแบบภาพ
2. ข้อตกลงเบื้องต้นว่า Lemur search toolkit สามารถสกัดคำในเว็บไซต์ได้ถูกต้องและแม่นยำสูง
3. เปรียบเทียบคำที่สกัดได้จากรูปภาพของทั้งหน้าเว็บเพจ
4. ใช้ตัวชี้วัด Precision, Recall, และ F-measure

5.3 สมมติฐานงานวิจัย (Hypotheses)

1. การสกัดคำโดยวิธีการของเราให้ความถูกต้องเมื่อเปรียบเทียบกับ Lemur search toolkit
2. จำนวนคำในภาพไม่มีผลต่อวิธีการสกัดคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4 ข้อตกลงเบื้องต้น (Assumption)

ในการทดลองของงานวิจัยนี้ เราใช้ผลการสกัดคำจากเครื่องมือการกระทำ (Lemur Search Toolkit) เป็นวิธีบรรทัดฐาน (baseline approach) เพื่อเปรียบเทียบโดยตั้งข้อตกลงเบื้องต้นว่า Lemur Search Toolkit นั้นถูกต้องทั้งหมด

5.5 ข้อมูลสำหรับการทดลอง

เราใช้ Clueweb09 category B [27] ซึ่งเป็นข้อมูลที่ถูกสร้างขึ้นมาเพื่อสนับสนุนการทำงานวิจัยเกี่ยวกับการค้นคืนสารสนเทศ (information retrieval) ประกอบเว็บเพจประมาณ 1,000 ล้านเว็บเพจ รองรับ 10 ภาษา ซึ่งเก็บในช่วงเดือนมกราคมถึงเดือนกุมภาพันธ์ ปี 2009 ชุดข้อมูลนี้ถูกนำไปใช้ในหลายแขนงในงานด้านการค้นคืนสารสนเทศ (Text Retrieval Conference TREC) [28] เนื่องจากขอบเขตของเราสนใจเฉพาะเว็บเพจภาษาอังกฤษ เราจึงเลือกใช้ Clueweb09 category B เพราะประกอบด้วยเว็บเพจเฉพาะภาษาอังกฤษเท่านั้น เราเลือกข้อมูลมา 1 ชุดจาก category B ซึ่งประกอบด้วยเว็บเพจประมาณ 30,000 เว็บเพจ เพื่อทำการทดลอง

5.6 Lemur search toolkit

Lemur search toolkit คือเครื่องมือที่ช่วยอำนวยความสะดวกในการเข้าถึงข้อมูลที่ถูกบีบอัดแบบพิเศษเนื่องจากข้อมูลมีขนาดใหญ่ Lemur search toolkit เป็นซอสโค้ด (source code) ซึ่งง่ายในการช่วยทำดัชนี (indexing) และดึงข้อมูลได้อย่างมีประสิทธิภาพ เราจึงใช้ Lemur search toolkit ในการทำดัชนีข้อมูล Clueweb09 category B เพื่อนำข้อมูลไปเปรียบเทียบกับการสกัดคำจากภาพที่สร้างจากเว็บเพจ ตารางที่ 5.1 แสดงจำนวนคำของเว็บเพจที่สุ่มขึ้นมาจาก 1,000 เว็บเพจเพื่อใช้ในการทดลอง ซึ่งได้หลังจากการทำดัชนีโดย Lemur search toolkit

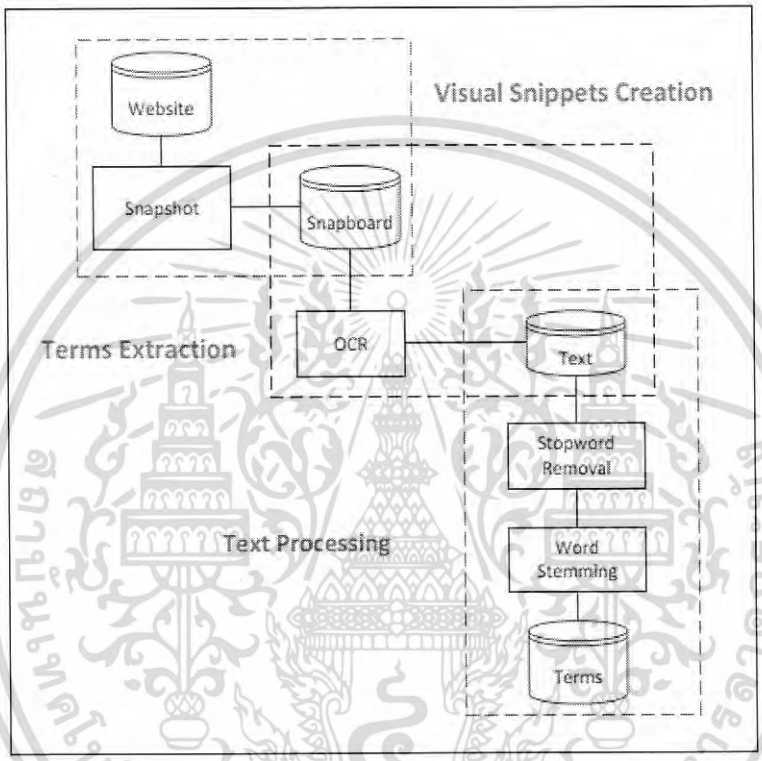
ตารางที่ 5.1 แสดงข้อมูลหลังทำดัชนีโดย Lemur search toolkit เพื่อใช้ในการทดลอง

Name	Web pages	Terms	Unique Terms
Clueweb09	1,000	1,307,646	439,227

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.7 การดึงคำจากรูปภาพ

รูปที่ 5.1 วิธีการสกัดคำจากรูปภาพหน้าเว็บเพจเบื้องต้น เพื่อเป็นแนวทางในการประยุกต์ใช้สร้างคำแนะนำสำหรับการค้นหาในอนาคต โดยแบ่งออกเป็น 3 ขั้นตอนดังนี้คือ 1) การสร้างรูปภาพเว็บเพจสำหรับการทดลอง 2) การดึงคำจากรูปภาพ และ 3) การประมวลผลคำที่ดึงได้จากภาพ มีรายละเอียดดังนี้

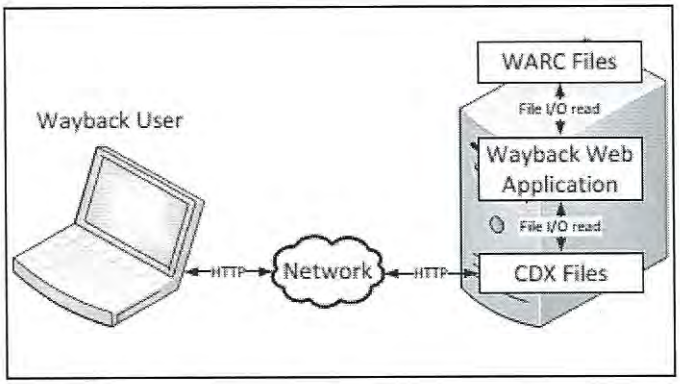


รูปที่ 5.1 วิธีการสกัดคำจากรูปภาพหน้าเว็บเพจเบื้องต้น

5.7.1 การสร้างรูปภาพเว็บเพจสำหรับการทดลอง

เนื่องจากข้อมูลใน Clueweb09 ถูกเก็บในรูปไฟล์นามสกุล WARC [29] ในการเข้าถึงข้อมูลประเภทนี้ เราใช้ Wayback tool [30] ในการทำดัชนี หลังจากทำดัชนีข้อมูลแล้ว เราจะได้ไฟล์ข้อมูลนามสกุล CDX ซึ่งเป็นไฟล์ข้อมูลที่ทำดัชนีแล้ว เราสามารถเข้าถึงข้อมูลในไฟล์นามสกุล CDX ได้ลิงค์แต่ละเว็บเพจด้วย (URL) หลังจากนั้นเราทำการสร้างรูปภาพของแต่ละเว็บเพจ โดยใช้ Python library ชื่อ Webkit2png [31] ซึ่งเป็นการนำลิงค์ของแต่ละเว็บเพจมาสร้างเป็นรูปภาพของเว็บเพจ

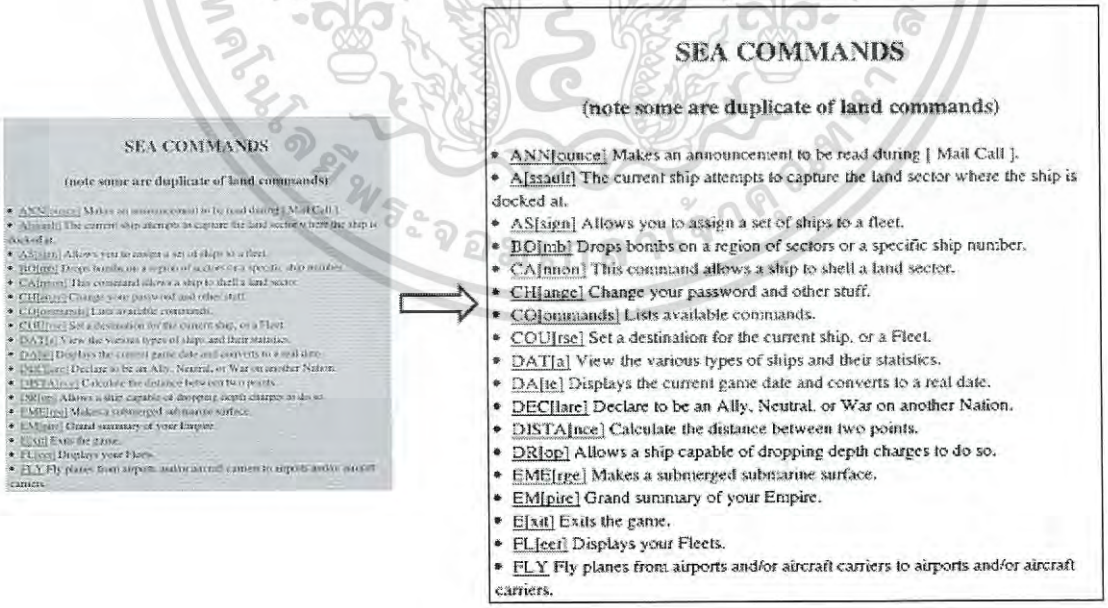
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 ภาพจำลองการทำงานของ Wayback tool

5.7.2 การดึงคำจากรูปภาพ

เพื่อที่จะใช้ประโยชน์จากภาพให้มากที่สุด เราทำการสกัดคำจากภาพ ด้วยกระบวนการ Optical character recognition (OCR) [32] โดยใช้เครื่องมือชื่อว่า Tesseract OCR [33] [34] เป็นชุดเครื่องมือโอเพนซอร์สที่ช่วยในการสกัดคำในภาพออกมาเป็นคำที่คอมพิวเตอร์สามารถอ่านได้และรองรับหลายภาษา หลังจากทำการทดลองสกัดคำ เราพบว่าภาพที่นำมาใช้ยังให้ผลของคำที่ไม่น่าพึงพอใจ เราจึงทำการปรับปรุงภาพก่อนกระบวนการสกัดคำ โดยขยายภาพให้มีขนาดใหญ่ขึ้นจากเดิมสามเท่า และปรับสีของภาพให้เป็นภาพขาว-ดำ เราพบว่าให้ผลดีขึ้นจากเดิมมาก เราจึงทำการปรับภาพทุกภาพในการทดลอง ตัวอย่างการปรับภาพแสดงในรูปที่ 5.3



รูปที่ 5.3 การขยายภาพให้ใหญ่ขึ้นจากเดิมสามเท่าและปรับภาพให้เป็นภาพขาว-ดำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

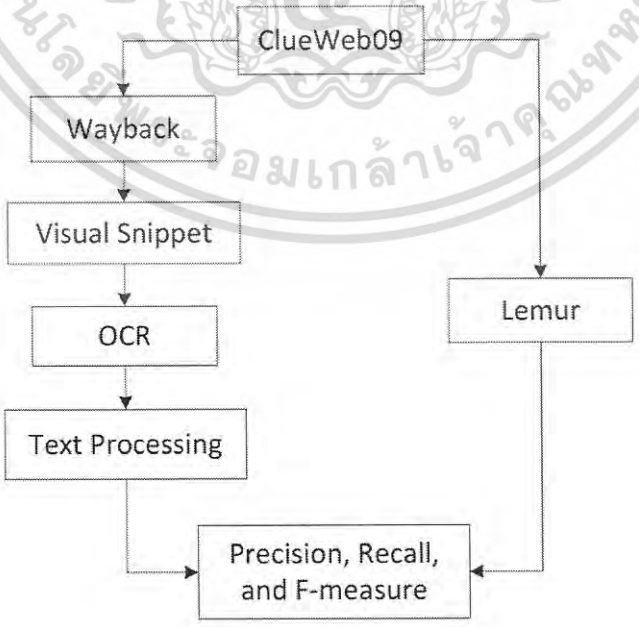
5.7.3 การประมวลผลคำที่ดึงได้จากภาพเบื้องต้น

หลังจากได้คำทั้งหมดแล้ว ขั้นตอนต่อไปเราทำการตัดคำที่เกิดขึ้นบ่อยและไม่สำคัญเรียก "Stopword" เช่น a, is, on, และ the เป็นต้น จากนั้น เราทำการแปลงคำให้อยู่ในรูปคำปกติเรียกว่า "Word stemming" เช่น processes, processed และ process สามารถแปลงเป็น process เป็นต้น

5.8 ขั้นตอนการทดลอง

5.8.1 การตรวจสอบการสกัดคำ

ในการการตรวจสอบการสกัดคำ เราใช้การเปรียบเทียบการสกัดคำของเรากับเครื่องมือชื่อ Lemur search toolkit ขั้นตอนในการทำการทดลองแสดงในรูปที่ 5.4 โดยสนใจเฉพาะเว็บเพจภาษาอังกฤษจาก Clueweb09 เราใช้ Wayback ในการเข้าถึงข้อมูล โดยเราสุ่มเว็บเพจมา 1,000 เว็บเพจ เราพบว่าเว็บเพจส่วนใหญ่มีจำนวนคำอยู่ในช่วง 100-400 คำ เพื่อที่จะตรวจสอบว่าจำนวนคำไม่มีผลต่อการสกัดคำ เราได้แบ่งกลุ่มเว็บเพจตามจำนวนคำ กลุ่มละ 250 เว็บเพจ ออกเป็น 4 กลุ่ม ดังนี้ 1) 50-200 คำ 2) 201-400 คำ 3) 401-600 4) มากกว่า 600 คำขึ้นไป เราใช้ Wenkit2png ในการสร้างรูปภาพจากเว็บเพจที่ได้ หลังจากนั้น ใช้เครื่องมือTesseract OCR ในการสกัดคำออกจากรูปภาพ และใช้การประมวลผลคำได้แก่ Stopword, Word stemming สุดท้ายเรานำคำที่ได้มาเปรียบเทียบกับคำที่ดึงจาก Lemur search toolkit โดยใช้ตัวชี้วัดคือ Precision, Recall และ F-measure



รูปที่ 5.4 ภาพจำลองขั้นตอนในการทำทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

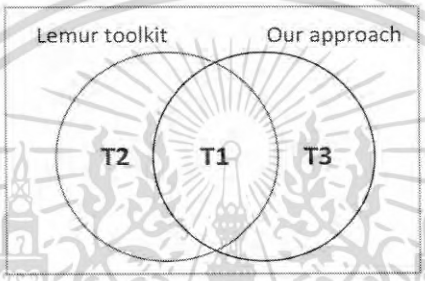
5.8.2 ตัวชี้วัด

Precision, Recall และ F-measure เป็นตัวชี้วัดพื้นฐานในการวัดประสิทธิภาพระบบค้นหาข้อมูล เราประยุกต์ Precision, Recall และ F-measure ในการเปรียบเทียบความถูกต้องในการสกัดคำจากภาพกับ Lemur search toolkit[1] เราแบ่งกลุ่มคำในการทดลองออกเป็น 3 กลุ่ม แสดงในรูปที่ 5.5 ดังนี้

T1 คือ กลุ่มของคำที่สกัดได้จากภาพ ซึ่งตรงกับคำที่ดึงได้จาก Lemur search toolkit

T2 คือ กลุ่มของคำที่ดึงได้จาก Lemur search toolkit แต่ไม่ตรงกับคำที่สกัดได้จากภาพ

T3 คือ กลุ่มของคำที่สกัดได้จากภาพ แต่ไม่ตรงกับคำที่ดึงได้จาก Lemur search toolkit



รูปที่ 5.5 ภาพจำลองเซตของคำ T1 T2 และ T3 สำหรับการทดลอง

1. **Precision*** (P^*) คือ อัตราส่วนของจำนวนคำที่สกัดออกจากภาพที่ตรงกับ Lemur search toolkit (N_{T1}) ต่อจำนวนคำทั้งหมดที่สกัดจากภาพ ($N_{T1}+N_{T3}$)

$$P^* = \frac{N_{T1}}{N_{T1}+N_{T3}} \tag{5.1}$$

2. **Recall*** (R^*) คือ อัตราส่วนของจำนวนคำที่สกัดออกจากภาพที่ตรงกับ Lemur search toolkit (N_{T1}) ต่อจำนวนคำทั้งหมดที่ดึงได้จาก Lemur search toolkit ($N_{T1}+N_{T2}$)

$$R^* = \frac{N_{T1}}{N_{T1}+N_{T2}} \tag{5.2}$$

3. **F-measure*** (F^*) คือ อัตราส่วนประสิทธิภาพโดยรวม คำนวณจากค่าความแม่นยำ และค่าระลึก

$$F^* = \frac{2 \times P^* \times R^*}{P^* + R^*} \tag{5.3}$$

[1] ตัวชี้วัด P^*, R^*, และ F^* ที่ถูกใช้เพื่อวัดประสิทธิภาพของการสกัดคำในการทดลองนี้ ถูกใช้ในการคำนวณภายใต้ข้อตกลงเบื้องต้นของการทดลองดังที่ชี้แจงไว้ในหัวข้อ 5.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.9 ผลการทดลอง

ตารางที่ 5.2 แสดงผลการทดลองพบว่า คำที่เราสกัดได้และตรงกับคำที่ได้จาก Lemur (T1) สูงถึง 357,721 คำ ซึ่งคิดเป็น 81.44 เปอร์เซ็นต์ของคำที่ Lemur ดึงได้ทั้งหมด และคำที่เราไม่สามารถสกัดออกมาได้ (T2) มีทั้งหมด 81,500 คำ คิดเป็น 18.56 เปอร์เซ็นต์ของคำที่ Lemur ดึงได้เท่านั้น ถึงแม้เราจะพบว่าคำที่เราสกัดออกมาไม่ตรงกับ Lemur (T3) มี 125,765 คำ คิดเป็น 28.63 เปอร์เซ็นต์ของคำที่ Lemur ดึงได้ แต่เราไม่สามารถตัดสินได้ว่า คำเหล่านี้ไม่มีประโยชน์ เพราะ Lemur เชื่อถือได้ว่าสามารถดึงคำที่ถูกต้องจากเว็บไซต์ แต่อาจจะดึงคำได้ไม่หมดเว็บไซต์ก็ได้ นอกจากนี้เรายังพบว่าวิธีการสกัดคำของเราให้ค่า P^*, R* และ F^* สูงคิดเป็น 73.9, 81.4 และ 77.5 เปอร์เซ็นต์ตามลำดับ ซึ่งทำให้สามารถตอบสมมติฐานที่ 1) ได้ว่า การสกัดคำจากรูปเว็บไซต์ให้ความถูกต้องใกล้เคียงกับการดึงคำจากเว็บไซต์โดยใช้เครื่องมือ Lemur search toolkit

ตารางที่ 5.2 แสดงผลการสกัดคำจากรูปภาพของ 1,000 เว็บเพจ

T1	357,721 terms (81.44%)
T2	81,506 terms (18.56%)
T3	125,765 terms (28.63%)
P^*	0.739
R^*	0.814
F^*	0.775

ค่า P^*, R^* และ F^* ของแต่ละกลุ่ม แยกตามจำนวนคำที่มีในเว็บเพจ แสดงในตารางที่ 5.3 พบว่ามีค่าสูงใกล้เคียงกันทุกช่วง เราจึงสามารถตอบสมมติฐานที่ 2) ได้ว่าจำนวนคำในภาพไม่มีผลต่อกระบวนการสกัดคำจากภาพของเรา

ตารางที่ 5.3 แสดงผลการสกัดคำจากรูปภาพแบ่งตามกลุ่มของจำนวนคำที่มีในเว็บไซต์

Group of terms	P^*	R^*	F^*
50-200 terms/pages	0.709	0.836	0.767
201-400 terms/pages	0.719	0.837	0.774
401-600 terms/pages	0.717	0.828	0.768
>600 terms/pages	0.767	0.796	0.781

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

สรุปผลงานวิจัยและข้อเสนอแนะ

6.1 สรุปผลงานวิจัย

วิทยานิพนธ์เล่มนี้นำเสนอระบบค้นหาข้อมูลแบบร่วมมือกันชื่อ "โคสเปช" และเครื่องมือใหม่ที่ช่วยให้ผู้ใช้สื่อสารกันระหว่างค้นหาได้ดียิ่งขึ้นชื่อ "กระดานแสดงและแบ่งปันสรุปแบบภาพภายในกลุ่ม" ซึ่งเป็นการนำรูปภาพมาใช้แทนข้อความตัวอักษรในการแบ่งปันข้อมูลของเว็บไซต์ที่น่าสนใจที่ผู้ค้นหาตัดด้วยตนเอง เพื่อปรับปรุงระบบค้นหาข้อมูลแบบร่วมมือกันที่มีอยู่ในปัจจุบันและยังไม่เป็นที่นิยม

เพื่อให้มั่นใจว่า กระดานแสดงและแบ่งปันสรุปแบบภาพภายในกลุ่ม ช่วยผู้ใช้ในงานที่คนมักจะร่วมมือกันค้นหาจริง การศึกษากลุ่มผู้ใช้งานระบบโคสเปชจึงเริ่มขึ้น โดยการเปรียบเทียบระหว่าง 2 ระบบ คือ ระบบที่มีฟังก์ชันการตัดภาพสรุปของเว็บไซต์ (S1) และระบบที่มีฟังก์ชันแสดงความคิดเห็นของเว็บไซต์ (S2) เราเลือกใช้โจทย์ 4 ข้อที่มาจาก 4 หัวข้อที่คนมักค้นหาร่วมกันมากที่สุด เราใช้ผู้ทดลองจำนวน 32 คน และจำลองสถานการณ์การค้นหาข้อมูลออกเป็น 2 แบบคือ แบบค้นหาคนเดียว (IN) และค้นหาร่วมกันแบบต่างเวลากัน (AC) การทดลองถูกออกแบบตามหลัก Graeco-Latin Square โดยการหมุนเวียนลำดับการทำโจทย์และระบบที่ใช้เพื่อแก้ปัญหาเรื่องการเรียนรู้ของมนุษย์ซึ่งจะทำให้ผู้ทดลองทำโจทย์ข้อหลังได้ดีกว่าข้อแรกเสมอ เราเก็บข้อมูลทั้งเชิงปริมาณจากการเก็บจำนวนกิจกรรมที่ผู้ทดลองทำระหว่างค้นหา และเชิงคุณภาพจากแบบสอบถามความพึงพอใจ

ผลการทดลองจากการเก็บข้อมูลเชิงปริมาณพบว่า ผู้ทดลองใช้การตัดภาพสรุปมากกว่าการแสดงความคิดเห็นทั้งในการค้นหาแบบ IN และ AC นอกจากนี้เราพบว่าผู้ทดลองใช้คำค้นหาในระบบ S1 มากกว่า S2 ทั้งในการค้นหาแบบ AC และ IN แต่ไม่ได้หมายความว่าพวกเขาต้องใช้ความพยายามในการคิดคำในการค้นหาข้อมูลใน S1 มากกว่า S2 เนื่องจากผู้ทดลองได้ระบุว่าการคิดคำที่ใช้ในการค้นหาของโจทย์ทั้ง 4 ข้อนั้นไม่ยาก เรายังพบว่าผู้ทดลองพบเว็บไซต์ที่เกี่ยวข้องหรือตรงกับความต้องการใน S1 มากกว่า S2 เมื่อพวกเขาใช้เวลาเท่ากันในการค้นหาข้อมูล ซึ่งหมายความว่ารูปภาพอาจจะช่วยให้พวกเขาสื่อสารและตัดสินเว็บไซต์ที่เกี่ยวข้องได้เร็วกว่าการอ่านความคิดเห็นในเวลาที่จำกัด ซึ่งอาจทำให้พวกเขาประเมินเว็บไซต์ที่น่าจะเกี่ยวข้องและตรงกับความต้องการภายในกลุ่มต่ำเกินไป เพราะในการทดลองแบบ AC เรานับเว็บไซต์ที่เกี่ยวข้องของกลุ่มเฉพาะเว็บไซต์ที่ผู้ทดลองทั้ง 2 คนในกลุ่มตัดสินว่าเป็นเว็บไซต์ที่เกี่ยวข้องทั้งคู่ และสุดท้ายเราพบว่าเปอร์เซ็นต์ของเว็บไซต์ที่เกี่ยวข้องต่อเว็บไซต์ที่ผู้ทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เข้าดูทั้งหมดใน S1 มากกว่า S2 ซึ่งแสดงให้เห็นว่าผู้ทดลองมีโอกาสเจอเว็บไซต์ที่ตรงกับความต้องการในระบบ S1 มากกว่าระบบ S2 ในเวลาที่จำกัด นอกจากนี้เรายังมีการวิเคราะห์ข้อมูลแยกตามประเภทผู้ทดลองแบ่งเป็น 1) กลุ่มนักเรียน และ 2) กลุ่มนักศึกษา พบว่าฟังก์ชันการตัดสรุปแบบภาพมีประโยชน์อย่างมากในการค้นหาแบบร่วมมือในกลุ่มนักเรียน แต่ในกลุ่มนักศึกษาเราพบว่าทั้งการตัดสรุปแบบภาพและการแสดงความคิดเห็นมีประโยชน์ทั้งคู่

ผลการทดลองจากการเก็บข้อมูลเชิงคุณภาพ เราพบว่าผู้ทดลองชอบระบบ S1 มากกว่าระบบ S2 ในโจทย์ประเภท การหาข้อมูลวางแผนท่องเที่ยว การหาข้อมูลซื้อของที่มีราคาแพง และการหาข้อมูลแก้ปัญหาเชิงเทคนิค เป็นเพราะว่าสรุปแบบภาพมีประโยชน์ในการแบ่งปันข้อมูลประเภทที่ต้องการให้เห็นภาพชัดเจนและเข้าใจง่ายกว่าข้อความตัวอักษร เช่น สถานที่ท่องเที่ยว หรือภาพสินค้าที่ต้องการซื้อ เป็นต้น แต่ผู้ใช้ชอบระบบ S2 มากกว่าระบบ S1 ในโจทย์ประเภทการค้นหาข้อมูลทางวรรณกรรม อาจเป็นเพราะว่างานวรรณกรรมเป็นงานที่ใช้การอ่านเพื่อเข้าถึงข้อมูลมากกว่าการทำความเข้าใจจากรูปภาพ แต่โดยรวมหลังผู้ทดลองทำโจทย์ทั้ง 4 ข้อเสร็จ พบว่าผู้ทดลองชอบระบบ S1 ที่มีฟังก์ชันการตัดภาพสรุปของเว็บไซต์มากกว่าระบบ S2 ที่มีฟังก์ชันแสดงความคิดเห็นของเว็บไซต์

นอกจากนี้เรายังนำเสนอการใช้ประโยชน์จากสรุปแบบภาพที่ผู้ใช้ตัดภายในกลุ่ม โดยการสกัดคำจากรูปภาพเพื่อนำมาสร้างคำแนะนำสำหรับการค้นหา การตรวจสอบความถูกต้องของการสกัดคำจากรูปภาพเราเปรียบเทียบกับเครื่องมือชื่อ Lemur search toolkit ที่สามารถใช้ทำดัชนีข้อมูล Clueweb09 category B ซึ่งเป็นชุดข้อมูลเว็บไซต์ภาษาอังกฤษที่เราใช้ในการทดลอง เราประยุกต์ค่า Precision, Recall, และ F-measure มาใช้ในการวัดความถูกต้อง ปรากฏว่าได้ผลที่น่าพึงพอใจคิดเป็น 73.9, 81.4 และ 77.5 เปอร์เซ็นต์ตามลำดับ เพื่อให้มั่นใจว่าจำนวนคำไม่มีผลต่อการสกัดคำของเรา เราจึงทำการแบ่งกลุ่มของรูปภาพตามจำนวนคำที่ปรากฏในเว็บไซต์ออกเป็น 4 กลุ่มคือ 1) 50-200 คำ 2) 201-400 คำ 3) 401-600 และ 4) มากกว่า 600 คำขึ้นไป ผลปรากฏว่าได้ค่า Precision, Recall, และ F-measure ที่ค่อนข้างสูงใกล้เคียงกันทั้ง 4 กลุ่ม เราจึงสรุปได้ว่าจำนวนคำไม่มีผลต่อการสกัดคำของวิธีการของเรา

6.2 ข้อเสนอแนะ

แม้ผู้ทดลองส่วนใหญ่จะชอบฟังก์ชันการตัดภาพสรุปของเว็บไซต์มากกว่าฟังก์ชันการแสดงความคิดเห็นเพราะช่วยในการสื่อสารกับเพื่อนในการค้นหาข้อมูลได้ง่ายและชัดเจนกว่า แต่ก็มีผู้ทดลองบางส่วนเห็นว่าการแสดงความคิดเห็นนั้นก็มีประโยชน์ซึ่งทำให้ผู้ค้นหาให้ข้อมูลเพิ่มเติมนอกเหนือจากที่มีในเว็บไซต์และละเอียดกว่ารูปภาพ ในส่วนของข้อเสียเราพบว่าฟังก์ชันการตัดภาพสรุปของเว็บไซต์ยังมีความช้าในการสร้างรูปภาพ เนื่องจากปัญหาทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เทคนิคซึ่งสามารถปรับปรุงได้ในอนาคต ส่วนฟังก์ชันแสดงความคิดเห็นก็มีโอกาสช้าเช่นกัน ในการคิดและพิมพ์ข้อความแสดงความคิดเห็น ซึ่งข้อความตัวอักษรอาจทำให้มองไม่เห็นภาพ และยิ่งข้อความที่ยาวมาก ก็อาจจะทำให้อ่านยากหรือทำให้รู้สึกไม่อยากอ่านอีกด้วย เราจึงคิดว่าการรวมกันของทั้ง 2 ฟังก์ชันในระบบค้นหาข้อมูลแบบร่วมมือน่าจะช่วยเพิ่มประสิทธิภาพในการค้นหาร่วมกันมากขึ้น นอกจากนี้มีผู้ทดลองบางส่วนแนะนำให้นำเสียงเข้ามาช่วยในการสื่อสารอีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] M. R. Morris, "A survey of collaborative web search practices," in *Proceedings of Computer Human Interaction.* CHI 2008, pp. 1657–1660. Florence, Italy.

[2] H. Kruajirayu, A. Tangsomboon, and T. Leelanupab, "การพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม," in *Proceedings of the National 6th Conference on Information Technology.* NCIT 2014, Thailand.

[3] H. Kruajirayu, A. Tangsomboon, and T. Leelanupab, "CoZpace: A proposal for collaborative web search for sharing search records and interactions," in *the 3rd ICT International Student Project Conference.* ICT-ISPC 2014 Third ICT International, Thailand.

[4] T. Leelanupab, H. Kruajirayu, and N. Kanungsukkasem, "Snapboard: A Shared Space of Visual Snippets. A Study in Individual and Asynchronous Collaborative Web Search." in *Proceedings of the Asia Information Retrieval Societies Conference.* AIRS 2015, Brisbane, Queensland, Australia.

[5] H. Kruajirayu, T. Leelanupab, "Extracting visual snippets for query suggestion in collaborative web search," in *Proceedings of the 20th International Symposium on Artificial Life and Robotics.* AROB 2015, Beppu, Japan.

[6] Pia Borlund and Peter Ingwersen, "The Development of a Method for the Evaluation of Interactive Information Retrieval Systems," *Journal of Documentation*, 53:225–250, 1997. 30

[7] Cyril W. Cleverdon, Jack Mills, and Michael Keen, "Factors Determining the Performance of Indexing Systems," *Technical report, ASLIB Cranfield Project*, 1966.

[8] Ellen M. Voorhees, "TREC: Improving Information Access through Evaluation," *Bulletin of the American Society for Information Science and Technology*, 32(1):16–21, 2005.

[9] Peter Ingwersen and Kalervo Jarvelin, "The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)," *Springer-Verlag*, 2005.

[10] Stephen E. Robertson and Micheline M. Hancock-Beaulieu, "On the Evaluation of IR System," *Information Processing & Management*, 28(4):457 – 466, 1992.

[11] P. Borlund, "User-centered evaluation of information retrieval systems," *Information Retrieval: searching in the 21st Century.* pp. 21–37 (2009.)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

[12] G. Golovchinsky, J. Pickens, and M. Back, "A taxonomy of collaboration in online information seeking," in *Proceedings of the 1st International Workshop on Collaborative Information Retrieval*. JCDL 2008, Pittsburgh, USA.

[13] S. Amershi, and M. R. Morris, "Cosearch: A system for co-located collaborative web search," in *Proceedings of Computer Human Interaction*. CHI 2008. Florence, Italy.

[14] M. Halvey, D. Vallet, D. Hannah, Y. Feng, J. M. Jose, "An asynchronous collaborative search system for online video search," *Information processing & management*. 46(6), 733–748 (2010)

[15] Z. Yue, S. Han, and D. He, "A comparison of action transitions in individual and collaborative exploratory web search," in *Proceedings of the Asia Information Retrieval Societies Conference*. AIRS 2012. Tianjin, China.

[16] M. R. Morris, and E. Horvitz, "Searchtogether: An interface for collaborative web search," in *Proceedings of User Interface Software and Technology*. UIST 2007. Newport, USA.

[17] M. R. Morris, J. Lombardo, and D. Wigdor, "Wesearch: supporting collaborative search and sensemaking on a tabletop display," in *Proceedings of Computer-Supported Cooperative Work and Social Computing*, CSCW 2010. Savannah, USA.

[18] J. R. Perez, T. Leelanupab, and J. M. Jose, "Cofox: A synchronous collaborative browser," in *Proceedings of the Asia Information Retrieval Societies Conference*. AIRS 2012. Tianjin, China.

[19] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. Andre, C. Hu, "Visual snippets: Summarizing web pages for search and revisitation," in *Proceedings of Computer Human Interaction*. CHI 2009. Boston, USA.

[20] B. Jiao, L. Yang, J. Xu, and F. Wu, "Visual summarization of web pages," in *Proceedings of Special Interest Group on Information Retrieval*. SIGIR 2010. Geneva, Switzerland.

[21] A. Aula, R. M. Khan, Z. Guan, P. Fontes, P. Hong, "A comparison of visual and textual page previews in judging the helpfulness of web pages," in *Proceedings of International World Wide Web Conference*. WWW 2010. Raleigh, USA.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

[22] Z. J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang, "Visual query suggestion," in *Proceedings of the 17th ACM International Conference on Multimedia.* MM 2009. Beijing, China.

[23] J. Denney, J. Hamilton, T. C. Hoad, K. R. Mcdonald, and R. J. Qian, "Visual query suggestions," U.S. Patent Application, 12/163,655, filed June 27, 2008.

[24] J. Bian, Z. J. Zha, H. Zhang, Q. Tian, and T.-S. Chua, "Visual query attributes suggestion," in Proceedings of the 20th ACM International Conference on Multimedia. MM 2012, Nara, Japan.

[25] P. Ogilvie, and J. P. Callan, "Experiments using the lemur toolkit," in *TREC*, volume 10, pages 103–108, 2001.

[26] J. Allan, J. Callan, K. Collins-Thompson, B. Croft, F. Feng, D. Fisher, J. Lafferty, L. Larkey, T. N. Truong, P. Ogilvie, et al. "The lemur toolkit for language modeling and information retrieval," 2003.

[27] J. Callan, M. Hoy, C. Yoo, and L. Zhao. Clueweb09 data set, 2009.

[28] D. Harman, "Overview of the first trec conference," in *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval.* 1993.

[29] I. ISO. 28500: 2009, "Information and documentation-WARC file format," *International Organization for Standardization.* 2009.

[30] B. Tofel, "Wayback for accessing web archives," in *Proceedings of the 7th International Web Archiving Workshop.* 2007.

[31] P. Hammond, "Webkit2png," *www.paulhammond.org/webkit2png/*, 2013.

[32] S. Mori, H. Nishida, and H. Yamada, "Optical character recognition," John Wiley & Sons, Inc., 1999.

[33] R. Smith, "An overview of the tesseract OCR engine," in *Proceedings of the 9th International Conference on Document Analysis and Recognition.* ICDAR 2007. Curitiba, Brazil.

[34] GOOGLE, "Tesseract-ocr," *http://code.google.com/p/tesseract- ocr/*, 2012.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

ผลงานวิจัยที่ได้รับการตีพิมพ์

ภาคผนวกนี้ได้นำเสนอบทความฉบับสมบูรณ์ของผลงานวิจัยซึ่งได้รับการตอบรับและตีพิมพ์ลงในการประชุมวิชาการระดับนานาชาติ ซึ่งเป็นส่วนหนึ่งของการทำวิทยานิพนธ์เล่มนี้ โดยมีรายการดังต่อไปนี้

1. การพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม

   Hannarin Kruajirayu, Ake Tangsomboon, and Teerapong Leelanupab; in the National 6th Conference on Information Technology, NCIT 2014; Thailand.

2. CoZpace: A Proposal for Collaborative Web Search for Sharing Search Records and Interactions.

   Hannarin Kruajirayu, Ake Tangsomboon, and Teerapong Leelanupab; in the 3rd ICT International Student Project Conference, ICT-ISPC 2014; Thailand.

3. Extracting Visual Snippets for Query Suggestions in Collaborative Web Search.

   Hannarin Kruajirayu, and Teerapong Leelanupab; in Proceedings of the 20th International Symposium on Artificial Life and Robotics, AROB 2015; Beppu, Japan.

4. Snapboard: A Shared Space of Visual Snippets. A Study in Individual and Asynchronous Collaborative Web Search.

   Teerapong Leelanupab, **Hannarin Kruajirayu**, and Nont Kanungsukkasem; in Proceedings of the Asia Information Retrieval Societies Conference, AIRS 2015; Brisbane, Queensland, Australia.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประชุมวิชาการระดับชาติด้านเทคโนโลยีสารสนเทศ ครั้งที่ 6
The 6th National Conference on Information Technology

โดย : คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

©สงวนสิทธิ์ โดย คณะกรรมการจัดการประชุมวิชาการระดับชาติด้านเทคโนโลยีสารสนเทศ ครั้งที่ 6 : NCIT2014

การนำงานวิจัยไปเผยแพร่หรืออนุญาตให้ผู้อื่นทำซ้ำ หรือดัดแปลง หรือเผยแพร่ต่อสาธารณชน หรือกระทำการอื่นใดโดยมีค่าตอบแทนในเชิงธุรกิจ โดยไม่ได้รับความยินยอมเป็นลายลักษณ์อักษรจากคณะกรรมการจัดการประชุมวิชาการระดับชาติด้านเทคโนโลยีสารสนเทศ ครั้งที่ 6 : NCIT2014 ถือเป็นการละเมิดลิขสิทธิ์

ISBN: 978-974-671-630-7

จัดพิมพ์ : ครั้งที่ 1 กุมภาพันธ์ 2557 จำนวน 50 เล่ม
พิมพ์ที่ : โรงพิมพ์มหาวิทยาลัยธุรกิจบัณฑิตย์
โทรศัพท์ 7300-2954-0 ต่อ 277 โทรสาร 0-2954-7908
ระบบการพิมพ์ : ดิจิทัล
ปก : กระดาษอาร์ตการ์ด น้ำหนัก 260 แกรมต่อตารางเมตร
เนื้อใน : ปอนด์ 70 แกรมต่อตารางเมตร
ขนาดรูปเล่ม : A4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประชุมวิชาการระดับประเทศด้านเทคโนโลยีสารสนเทศ (National Conference on Information Technology: NCIT) ครั้งที่ 6

การพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม

หาญนรินทร์ เครือจิรายุส์ เอก ตั้งสมบูรณ์ และ ธีรพงศ์ ลีลานุภาพ

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Emails: hannarin_k@hotmail.com, ake@mercigod.com, teerapong@it.kmitl.ac.th

บทคัดย่อ

เสิร์ชเอนจิน (Search Engine) ถูกใช้งานเพื่อการค้นหาข้อมูลอย่างแพร่หลาย ระบบการค้นคืนในปัจจุบันถูกออกแบบมาสำหรับการใช้งานสำหรับผู้ใช้เพียงคนเดียว แต่ในงานบางประเภท ผู้ใช้จำเป็นที่จะต้องค้นหาข้อมูลร่วมกันมากกว่าหนึ่งคนหรือร่วมกันเป็นทีม ทั้งนี้เนื่องจากข้อมูลที่ต้องการค้นหานั้นอาจมีเป็นจำนวนมากหรือยากแก่การค้นหาโดยเพียงคนเดียว การที่ผู้ใช้แต่ละคนทำการค้นหาข้อมูลร่วมกันแต่แยกกันทำ อาจทำให้ข้อมูลที่ค้นหาได้นั้นมีความซ้ำซ้อนหรืออาจไม่เกี่ยวข้องกับสิ่งที่ต้องการ และทำให้เสียเวลาเพิ่มเติมในการค้นหา จึงเกิดแนวคิดในการพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม (Collaborative Search) บทความวิจัยนี้นำเสนอการพัฒนาระบบการร่วมมือค้นหาข้อมูลแบบกลุ่ม ชื่อ CoZpace โดยระบบถูกพัฒนาโดยใช้เทคโนโลยี Web socket ในการติดต่อสื่อสารระหว่างเว็บเซิร์ฟเวอร์กับไคลเอนต์ ซึ่งเป็นรูปแบบการร่วมมือในเวลาเดียวกัน (Synchronous Collaboration) และมีการเก็บข้อมูลประวัติการใช้งานระบบของผู้ใช้ภายในกลุ่มไว้ในฐานข้อมูลเพื่อนำมาแสดง เช่น คำค้นหาและการคลิ๊กของผู้ใช้บนผลการค้นหา ทั้งนี้เพื่อรองรับการร่วมมือแบบต่างเวลากัน (Asynchronous Collaboration) ซึ่งทำให้เกิดการตระหนัก (Awareness) ถึงความร่วมมือระหว่างผู้ร่วมค้นหา

คำสำคัญ - โคเสปซ (CoZpace); การร่วมมือค้นหาข้อมูลแบบกลุ่ม (Collaborative Search); การค้นคืนสารสนเทศ (Information Retrieval); กระดานแสดงและแชร์ข้อมูลระหว่างกลุ่ม (Snapboard); การร่วมมือแบบเปิดเผย (Explicit Collaboration); การร่วมมือในเวลาเดียวกัน (Synchronous Collaboration); การร่วมมือต่างเวลากัน (Asynchronous Collaboration)

1. บทนำ

เสิร์ชเอนจินเป็นเครื่องมือที่ช่วยในการค้นหาข้อมูลไม่ว่าจะเป็นข้อมูลที่เป็นข้อความหรือรูปภาพ โดยการประยุกต์ใช้ศาสตร์ทางด้านการค้นคืนสารสนเทศ (Information Retrieval) ในการค้นคืนข้อมูลของกลุ่มเอกสารขนาดใหญ่ ซึ่งเกี่ยวกับการจัดเก็บข้อมูลที่ตอบสนองความต้องการของผู้ใช้และจัดอันดับข้อมูลตามความเกี่ยวข้องต่อความต้องการข้อมูล (Information needs) ของผู้ใช้

ในปัจจุบันนั้น หนึ่งในงานวิจัยด้านการค้นคืนสารสนเทศ คือ การร่วมมือค้นหาข้อมูลแบบกลุ่ม (Collaborative Search) ซึ่งเป็นการศึกษาเพื่อสนับสนุนการค้นหาข้อมูลร่วมกันโดยผู้ใช้มากกว่าหนึ่งคน ทำให้ช่วยในการค้นหาข้อมูลที่ตอบสนองความต้องการ ลดความซับซ้อนของข้อมูลที่ค้นหาร่วมกัน และอำนวยความสะดวกในการสื่อสารเพื่อวางแผนการค้นหาภายในกลุ่ม ซึ่งประโยชน์ของระบบการร่วมมือค้นหาข้อมูลแบบกลุ่มนั้นจะเป็นเครื่องมือที่ทำให้ผู้ใช้หลายๆคนสามารถค้นหาข้อมูลเดียวกันพร้อมๆกันได้ ทั้งต่างสถานที่และต่างเวลา โดยรูปแบบความร่วมมือในสาขาการค้นคืนสารสนเทศสามารถแบ่งออกเป็น 2 มิติ ได้แก่ *i)* รูปแบบการรับรู้ในการร่วมมือค้นหา *ii)* รูปแบบเวลาในการร่วมมือค้นหา

โดยในงานวิจัยของเรานั้นทำการนำเสนอการพัฒนาระบบค้นหาข้อมูลที่มีชื่อว่า CoZpace (โคเสปซ) ซึ่งมีฟังก์ชันต่างๆที่สามารถช่วยให้ผู้ใช้สามารถร่วมมือกันค้นหาได้อย่างสะดวก เช่น การสนทนาออนไลน์, การแสดงเว็บที่ถูกเข้าไปแล้ว, การแสดงความคิดเห็นแก่เว็บต่างๆ และ Snapboard (กระดานแสดงและแชร์ข้อมูลระหว่างกลุ่ม) ซึ่งทำให้ผู้ใช้สามารถตัดส่วนของหน้าเว็บส่วนหนึ่งเพื่อแสดงให้กับผู้ร่วมค้นหาเห็น และแสดงความคิดเห็นได้ โดยจะมีการประเมินระบบในเชิงคุณภาพ เพื่อแสดงให้เห็นว่าระบบช่วยให้สามารถค้นหาข้อมูลได้อย่างรวดเร็ว และตอบสนองความต้องการของผู้ใช้ได้

2. ทฤษฎีที่เกี่ยวข้องกับงานวิจัย

งานด้านการร่วมมือค้นหาข้อมูลแบบกลุ่ม มีทฤษฎีที่เกี่ยวข้องในด้านของการทำงานที่เป็นการสื่อสารกันระหว่างผู้ร่วมค้นหา รวมถึงการทำงานที่สื่อสารในเชิงเวลา ซึ่งจะอธิบายในอันดับถัดไป

2.1 รูปแบบการรับรู้ในการร่วมมือค้นหา

Implicit Collaboration (การร่วมมือโดยนัย) นั้นคือผู้ใช้จะไม่ทำงานร่วมกันโดยตรง โดยระบบจะทำการวิเคราะห์และกรองข้อมูลจากผู้ใช้หนึ่งคน เพื่อนำมาปรับปรุงผลการค้นหาสำหรับผู้ใช้อีกคนหนึ่ง โดยผู้ใช้ทั้งคู่นั้นจะไม่รู้ถึงการทำงานนี้ของระบบ โดยระบบจะนำข้อมูลต่างๆ ที่ได้จากผู้ใช้ เช่น กิจกรรมต่างๆ ที่ผู้ใช้มีต่อระบบ หรือ ระยะเวลาที่ผู้ใช้อยู่ในเว็บไซต์ต่างๆ มาทำเหมืองข้อมูลการใช้งานเพื่อนำมาปรับปรุงผลการค้นหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Explicit Collaboration (การร่วมมือแบบเปิดเผย) นั้นคือผู้ใช้สามารถแบ่งข้อมูลและทำงานร่วมกันไปสู่เป้าหมายที่ต้องการ เช่น instant message, คีย์เวิร์ด [1] โดยระบบของเรานั้นมีการใช้การค้นหาข้อมูลร่วมกันแบบเปิดเผย คือ ผู้ใช้จะร่วมมือกันโดยผ่านเครื่องมือของระบบ เช่น instant message, ปุ่ม relevant ซึ่งเป็นปุ่มที่ใช้บอกว่าเว็บที่เราดูอยู่นั้นเป็นเว็บที่เกี่ยวข้องกับความต้องการ

2.2 รูปแบบการเวลาในการร่วมมือค้นหา

Synchronous Collaboration (การร่วมมือในเวลาเดียวกัน) นั้นเป็นการทำงานร่วมกันแบบ เรียลไทม์คือ ผู้ร่วมค้นหาทุกคนทำงานด้วยกันไปพร้อมๆ กัน ไม่ว่าจะอยู่สถานที่เดียวกัน หรือ ต่างสถานที่กันก็ตาม ซึ่งเรียกได้ทั้ง 2 แบบคือ (same-time, same-place) (เวลาเดียวกัน สถานที่เดียวกัน) หรือ (same-time, different-places) (เวลาเดียวกัน, ต่างสถานที่กัน) ทำให้มีการทำงานร่วมกันที่เร็ว ในการวางแผน การแก้ปัญหา การจัดตารางเวลาและการตัดสินใจ ตัวอย่างเช่น การสนทนาด้วย instant message, การประชุมทางไกลด้วยวีดีโอ

Asynchronous collaboration (การร่วมมือต่างเวลากัน) นั้นเป็นการทำงานร่วมกันที่ระบบจะทำการบันทึกข้อมูล การทำงานร่วมกันเช่น ประวัติการสนทนา ประวัติคำที่ใช้ค้นหา โดยสามารถนำมาแสดงผลได้ในภายหลัง ซึ่งเรียกว่าแบบ (different-times, different-places) (เวลาใดๆ สถานที่ใดๆ) ทำให้ผู้ร่วมค้นหาสามารถที่จะมีส่วนร่วมเมื่อใดก็ได้อย่างอิสระ [2]

3. งานวิจัยที่เกี่ยวข้อง

ตัวอย่างงานวิจัยที่เกี่ยวข้องเพื่อใช้ในการศึกษาการร่วมมือค้นหาข้อมูลแบบกลุ่มมีดังนี้ คือ

3.1 A Comparison of Action Transitions in Individual and Collaborative Exploratory Web Search

บทความนี้พูดถึงการค้นหาร่วมกันที่มีหลายแบบ และพฤติกรรมของผู้ใช้งาน ในด้านการทำงานร่วมกัน โดยศึกษาจากรูปแบบการทำงานที่แตกต่างกัน บทความนี้นำเสนอการเปรียบเทียบความแตกต่างของวิธีการค้นหา 3 เงื่อนไข

- ผู้ใช้ค้นหาร่วมกัน แบบมีการสื่อสารที่ชัดเจน
- ผู้ใช้ค้นหาร่วมกัน แบบมีการสื่อสารที่ไม่ชัดเจน
- ผู้ใช้ค้นหาคนเดียว ไม่สื่อสารกัน

การค้นหาแบบ Explicit Collaboration ระหว่างผู้ใช้ ยังส่งเสริมให้เกิด Implicit Collaboration การศึกษานี้ให้คำแนะนำ เกี่ยวกับระยะเวลาในการค้นหาและกิจกรรมที่ระบบการร่วมมือค้นหาข้อมูลแบบกลุ่มควรสนับสนุน [3]

3.2 Aggregated Search Interface Preferences in Multi-Session Search Tasks

Aggregated search interfaces จะแสดงผลการค้นหา ในแบบให้เห็นภาพรวมของผลลัพธ์การค้นหาจากหลายแหล่ง มี 3 แบบคือ

- A tabbed display
- A blended display and a blended display
- A similar functionality

งานก่อนหน้านี้ศึกษาเกี่ยวกับ Vertical search (การค้นหาเฉพาะหัวข้อหรือประเภทข้อมูล) มีผลกระทบต่อพฤติกรรมการค้นหาของผู้ใช้หรืออิทธิพลของ tabbed และ blended interfaces กับพฤติกรรมการค้นหาของผู้ใช้แบบ single-session search task แต่งานวิจัยนี้ จะศึกษาเกี่ยวกับอิทธิพลของ tabbed และ blended interfaces กับพฤติกรรมการค้นหาของผู้ใช้แบบ multi-session search task [4]

3.3 Collaborative search revisited

บทความนี้ พูดถึงความก้าวหน้าและความรู้ที่เกี่ยวกับ Collaborative web search โดยเสนอข้อมูลการสำรวจ ผู้ใช้ที่แตกต่างกัน พบว่าการร่วมมือค้นหาข้อมูลแบบกลุ่มมีความแพร่หลายมากขึ้น (มีการพัฒนาเทคโนโลยีที่เพิ่มมากขึ้น เช่น smartphone และ social networking sites) การเปรียบเทียบกับงานวิจัยก่อนหน้า พบการเปลี่ยนแปลงของความแพร่หลายในการใช้ เครื่องมือรวมกลุ่ม และ เทคโนโลยี

แม้จะมีการเพิ่มขึ้นของ เครื่องมือที่ใช้ค้นหาร่วมกัน (Coagmento, Search Together และ ResultsSpace) แต่ในกลุ่มสำรวจไม่มีการใช้เทคนิคดังกล่าว แต่กลับพูดถึงเทคโนโลยีง่ายๆ ที่ใช้ในชีวิตประจำวัน เช่น e-mail, texting, instant messaging, phone calls and social networking [5]

4. แนวทางการดำเนินงาน

4.1 การศึกษาและรวบรวมข้อมูลระบบ

1. Coagmento

เป็น plug-in บน Firefox ช่วยให้สามารถค้นหาข้อมูลร่วมกันได้ โดยการเพิ่มสมาชิก และระบบของ Coagmento แบ่งเป็น 2 ส่วนหลัก คือ

- Toolbar ประกอบด้วย ปุ่มที่สามารถบันทึกหรือลบเว็บเพจ และให้คำอธิบายประกอบเว็บเพจ ปุ่มสำหรับดูสถิติในแต่ละเว็บเพจ เช่น จำนวนครั้งที่เข้าไปดูเว็บเพจ ผู้ใช้สามารถกดดูสถิติที่ได้บันทึกเอาไว้ จะได้รับข้อมูลเพิ่มเติม และปุ่มสำหรับดูสถิติในแต่ละงาน เช่น จำนวนหน้าเว็บเพจที่เข้าชม
- Sidebar ประกอบด้วยด้วยช่องสนทนา ใช้ช่องสนทนาประกอบด้วย 3 แถบเกี่ยวกับประวัติของการค้นหา (queries) เว็บเพจที่ถูกบันทึก (pages) และข้อความสั้นๆ (snippets) ทุกคนในกลุ่มสามารถเข้าถึงข้อมูล ที่สนใจโดยการคลิก ตัวอย่างเช่น คลิกที่แถบ queries จะแสดงคำที่ค้นหาและผู้ที่ค้นหาด้วยคำนั้นทั้งหมดของกลุ่ม ในหน้าเบราว์เซอร์หลัก [6]

2. ResultsSpace

เป็นระบบที่เพิ่มประสิทธิภาพด้าน awareness ให้ผู้ค้นหาข้อมูลโดยประกอบด้วย

- Query box ใส่คีย์เวิร์ดเพื่อการค้นหา
- Result displays จะแสดง link ตรงกลางหน้าโดยข้างซ้ายของแต่ละ link จะมีการให้ผู้ใช้สามารถให้ rate แก่ link นั้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Document view page จะมีการแสดงข้อมูลในเว็บนั้นๆ รวมถึงมีการให้ rating ภายใน links นั้นๆ เหมือนใน result display รวมถึงมีปุ่ม back เพื่อให้สามารถกลับไปที่ผลการค้นหาได้
- Queries histories ในส่วนทางด้านขวาจะมีการแสดง คีย์เวิร์ดที่ถูก ใช้ไปแล้ว โดยจะแสดง 5 อันดับสูงสุดของ คีย์เวิร์ดที่ถูกค้นหาและมีการแสดงแถบสี ที่แสดงอยู่ใต้คีย์เวิร์ดแต่ละคำ โดยจะสามารถระบุได้ว่า คีย์เวิร์ดนั้น มีกี่document ที่ relevant, non-relevant หรือ maybe [7]

3. Search Together

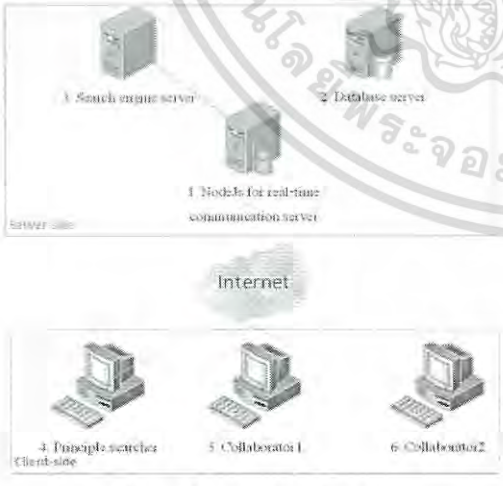
เป็นระบบที่สร้างขึ้นโดย Microsoft มี 3 ส่วนที่สำคัญคือ

- Awareness มีการแสดงสถานะของผู้ใช้ ถ้ารูปผู้ใช้ เป็นสีเทา แสดงว่า ไม่อยู่ในระบบแล้ว ในการค้นหา จะมีการบันทึกข้อมูลการค้นหา เพื่อลดความซ้ำซ้อนการใช้ คีย์เวิร์ดเดิม
- Division of Labor โดยเมื่อผู้ใช้เจอเว็บไซต์ที่อยากให้เพื่อนในทีมเห็น โดยจะสามารถเลือกกลุ่มของผู้ใช้ เพื่อทำการแนะนำ และ comments ได้
- Persistence เมื่อผู้ใช้มีการเข้าสู่ระบบเขาสามารถเห็นข้อมูลต่างๆ ที่ผู้ใช้คนอื่นได้ทำไว้ ในตอนที่เขา ออฟไลน์และเก็บข้อมูลต่างๆ ที่เขาทำไว้ล่าสุด เพื่อเป็นการเตือนให้รู้ว่าทำอะไรไปแล้วบ้าง[8]

4. CoFox: A Synchronous Collaborative Browser

เป็นระบบแบบ synchronous collaboration โดยผู้ใช้จะทำงานร่วมกัน 2 คน จะต้องทำงานในเวลาเดียวกัน โดยมีการทำ awareness มีการดูผ่านหน้าต่าง remote view ซึ่งจะแสดงหน้าต่างเบราเซอร์ของผู้ร่วมค้นหา ทำให้สามารถช่วยในการประเมิน เว็บเพจร่วมกันได้ โดยใช้ พื้นที่ร่วมกัน [9]

4.2 การออกแบบระบบ



รูปที่ 1 สถาปัตยกรรมของระบบ CoZpace

รูปที่ 1 แสดงระบบค้นคืนสารสนเทศสำหรับการค้นหาข้อมูลแบบกลุ่มความร่วมมือ และส่วนการทำงานต่างๆ ซึ่งสามารถแบ่งออกเป็น 5 ส่วนหลัก ดังต่อไปนี้

1. NodeJs real-time communication server เป็นส่วนที่ให้บริการการติดต่อสื่อสารแบบ "เรียลไทม์" เช่น ตัวอย่างที่เห็นได้ชัดเจนคือ สนทนาออนไลน์ และทำหน้าที่เป็นเว็บเซิร์ฟเวอร์
2. ฐานข้อมูล เป็นส่วนที่รวบรวมข้อมูลต่างๆ ที่เกี่ยวข้องกันเข้าไว้ด้วยกัน ในระบบฐานข้อมูลจะประกอบด้วยแฟ้มข้อมูล มีข้อมูลเกี่ยวข้องสัมพันธ์กันเข้าไว้ด้วยกันอย่างเป็นระบบ
3. Search engine server เป็นส่วนที่ใช้ในการรวบรวมเว็บเพจช่วยในการค้นหาข้อมูล โดยเฉพาะข้อมูลบนอินเทอร์เน็ต โดยครอบคลุมทั้งข้อความ รูปภาพ ภาพเคลื่อนไหว
4. Principle searcher เป็นคนที่ทำการสร้างโปรเจกต์เพื่อทำการค้นหาข้อมูล สามารถเพิ่ม ผู้ร่วมค้นหา เข้ามาในโปรเจกต์เพื่อร่วมกันในการค้นหาข้อมูล และเป็นเพียงคนเดียวที่มีสิทธิ์ในการปิดโปรเจกต์ที่ตัวเองสร้างขึ้น
5. ผู้ร่วมค้นหาเป็นคนที่ถูก principle searcher เพิ่มเข้ามาในโปรเจกต์เพื่อช่วยในการค้นหาข้อมูลร่วมกัน ซึ่งสามารถทำกิจกรรมต่างๆ ที่ระบบมีให้ได้เหมือนกับ principle searcher แตกต่างกันตรงที่ผู้ร่วมค้นหา ไม่มีสิทธิ์ในการปิดโปรเจกต์แต่สามารถที่จะออกโปรเจกต์เมื่อไม่ต้องการที่จะเข้าร่วมการค้นหาข้อมูลนั้นได้

4.3 การพัฒนาระบบ

1. การทำงานแบบเรียลไทม์

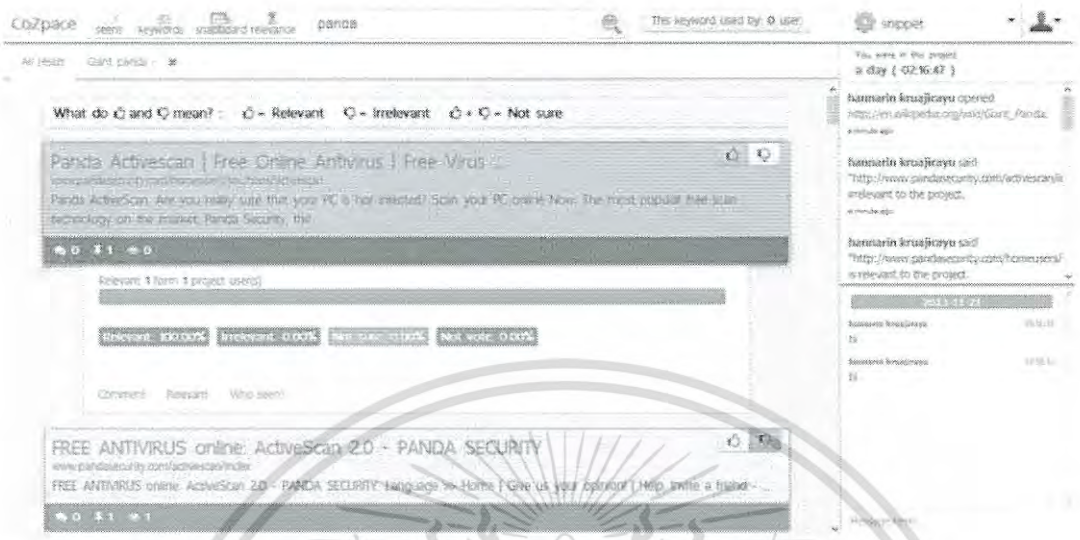
1.1 Web Sockets

เป็น API ต่อยอดแนวทางของเอแจ็กซ์ คือการส่งข้อมูลจากเซิฟเวอร์ไปไคลเอนต์ การส่งข้อมูลแบบ HTTP จะเปิดการเชื่อมต่อกับ เซิฟเวอร์เพื่อส่งข้อมูล แล้วตัดการเชื่อมต่อเมื่อใช้เสร็จ ดังนั้นการขอข้อมูลจากเซิฟเวอร์เป็นระยะจึงทำได้ยาก เพราะต้องดึงข้อมูลจาก เซิฟเวอร์เป็นระยะ ซึ่งเปลืองโหลดของเซิฟเวอร์โดยเฉพาะกรณีที่ต้องเปิดการเชื่อมต่อ http ค้างเอาไว้ Web sockets เกิดมาเพื่อแก้ปัญหานี้ โดยสร้างการเชื่อมต่อแบบ (เกือบ) ถาวรระหว่าง เซิฟเวอร์กับไคลเอนต์ เพื่อให้ทั้งสองฝั่งส่งข้อมูลกันได้ [10]

1.2 NodeJs

เป็น server side javascript ซึ่งใช้แนวคิดแบบ asynchronous event-driven model รองรับการทำงานพร้อมกัน เพื่อลด overhead ระหว่าง thread ซึ่งเหมาะกับระบบที่ต้องการทำงานแบบ "เรียลไทม์" การทำงานของโปรแกรมจะเป็นแบบ asynchronous หมายความว่า I/O จะไม่ถูกบล็อคการทำงาน เช่น การใช้งาน เอแจ็กซ์ เมื่อเริ่มทำงานก็จะ request ข้อมูลออกไประหว่างที่รอ response คำสั่งในบรรทัดถัดไปจะสามารถทำงานต่อได้ (ถ้าทำเป็นแบบ synchronous จะต้องรอให้มีการ response กลับมาก่อนจึงจะทำคำสั่งต่อไปได้) เมื่อมีการ response ข้อมูลกลับมา ก็จะมีการส่งอีเว้นท์บอกว่ามีการตอบกลับมาแล้ว ซึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2 Interface ของระบบ CoZpace แสดงผลการค้นหา

อีเว้นท์ตรงนี้เราสามารถเขียนโค้ดเพื่อจับ อีเว้นท์แล้วนำข้อมูลไปประมวลผลต่อ เช่น เอาไปแสดงผลหน้าเว็บเพจต่อได้ [11]

2. *ผลการค้นหา*

Bing search API เป็นบริการข้อมูลเกี่ยวกับการค้นหาหรือเสิร์ช โดยเป็นบริการจากเว็บไซต์ของ bing.com ที่เปิดบริการเสิร์ช แบบ API ที่ให้เราสามารถทำการนำผลลัพธ์ที่ได้จากค้นหาของเว็บไซต์ Bing มาใช้งานในด้านต่างๆ ได้ [12]

4.4 แผนการทดสอบ

เรามีการตั้งสมมติฐานเพื่อประเมินระบบ โดยจะทำการเก็บข้อมูลในเชิงคุณภาพ โดยจะใช้แบบสอบถามในการถามถึงความพึงพอใจ และประสิทธิภาพของระบบ CoZpace โดยจะใช้การให้คะแนนแบบ 5 point likert-scale questions เพื่อมาใช้ในการประเมิน และมีคำถามปลายเปิดเพื่อให้ผู้ใช้งานระบบ CoZpace ได้แสดงความคิดเห็นได้อย่างเต็มที่ โดยให้แบบสอบถามเมื่อผู้ใช้ได้ทำการทดลองเสร็จสิ้นแล้ว

1. สมมติฐานดังต่อไปนี้ได้คิดค้นขึ้นเพื่อประเมินระบบของเรา
   - 1.1. CoZpace ช่วยให้ผู้ใช้ ทำงานร่วมกันได้อย่างมีประสิทธิภาพ เมื่อทำงานร่วมกัน เมื่อค้นหาข้อมูลทั่วไป
   - 1.2. CoZpace ช่วยให้ผู้ใช้สามารถค้นหาข้อมูลที่เกี่ยวข้องได้มากขึ้นและเร็วกว่าเมื่อเทียบกับระบบพื้นฐาน
   - 1.3. การใช้ Snapboard ช่วยสนับสนุนทั้ง Synchronous และ Asynchronous Collaboration

2. คำถามแบบให้คะแนนมีดังนี้
   - 2.1 การแสดงกิจกรรมต่างๆของผู้ร่วมค้นหา ช่วยให้สามารถค้นหาได้รวดเร็วขึ้น
   - 2.2 การ relevant และแสดงความคิดเห็น ช่วยให้สามารถค้นหาได้รวดเร็วขึ้น
   - 2.3 การแสดงคีย์เวิร์ดที่ใช้แล้ว และ เว็บที่เข้าแล้ว ช่วยให้สามารถค้นหาได้รวดเร็วขึ้น
   - 2.4 การสนทนาผ่าน Instant message ช่วยให้สามารถค้นหาข้อมูลได้รวดเร็วขึ้น
   - 2.5 การใช้ Snapboard ช่วยให้สามารถทำงานร่วมกันได้ง่ายขึ้น

3. คำถามแบบปลายเปิดมีดังนี้
   - 3.1 ส่วนไหนของระบบ CoZpace ที่สามารถช่วยเพิ่มประสิทธิภาพของการค้นหา เพราะอะไร
   - 3.2 คิดว่าระบบ CoZpace ควรมีอะไรเพิ่มเติม เพื่อให้ผู้ใช้ช่วยกันค้นหาได้รวดเร็วขึ้น

5. ผลการดำเนินงาน

5.1 ผลการพัฒนาระบบ CoZpace

CoZpace เป็นระบบค้นหาข้อมูลเหมาะทั้งแบบผู้ใช้ค้นหาเพียงคนเดียว หรือแบบค้นหาร่วมกันเป็นทีม ซึ่งรองรับทั้ง direct และ indirect communication ระหว่างเพื่อนร่วมทีม ซึ่งในด้าน indirect communication ผู้ใช้สามารถดูประวัติการค้นหา ของเพื่อนร่วมทีมได้ สำหรับในด้าน direct communication พวกเขาสามารถใช้ instant message ที่มีให้ทางด้านขวามือ เพื่อการสนทนากับเพื่อนร่วมทีม จะมีหน้าหลักๆ 3 หน้าคือ *i)* หน้าจัดการโปรเจกต์ของระบบ *ii)* หน้าแสดงผลของเว็บไซต์ที่เข้าดู *iii)* หน้าแสดงผลการค้นหาทั้งหมด

จากรูปที่ 2 ระบบ CoZpace หน้าแสดงผลการค้นหาทั้งหมดประกอบด้วย 6 ส่วนที่สำคัญคือ

1. 4 ปุ่มด้านซ้ายบน เป็นส่วนที่ใช้แสดงสรุปข้อมูลต่างๆ โดยจะมีข้อมูลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

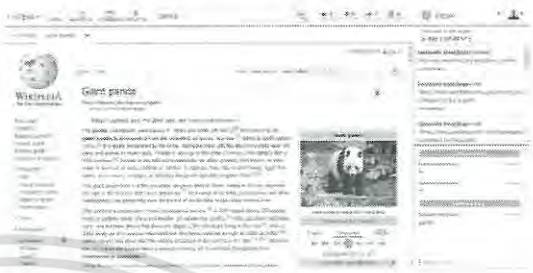
1.1 Seen จะแสดงเว็บไซต์ที่มีการเข้าดู ในโปรเจกต์ทั้งหมด
1.2 Keyword จะแสดงคีย์เวิร์ดทั้งหมดที่ใช้ค้นหา
1.3 Snapboard จะแสดงรูปทั้งหมดที่ได้ตัดเอาไว้
1.4 Relevant จะแสดงเว็บไซต์ที่มีการ Relevant, Non-relevant, Not Sure ทั้งหมดในโปรเจกต์ปัจจุบัน

2. ช่องตรงกลางเป็นส่วนไว้ใส่ คีย์เวิร์ดที่ใช้ค้นหา
3. ตรงกลางเป็นส่วนแสดงผลเว็บไซต์ เป็นผลลัพธ์ของการค้นหา จะมีปุ่มซ้ายบน ซึ่งคือ relevant และ non--relevant ไว้ให้ rate เว็บไซต์ โดยจะมี 3 ลักษณะดังนี้
   3.1 สีเขียว เมื่อผู้ใช้มีการกดปุ่ม Relevant
   3.2 สีแดง เมื่อผู้ใช้มีการกดปุ่ม Non-Relevant
   3.3 สีเหลือง เมื่อผู้ใช้มีการกดทั้งปุ่ม Relevant และ Non-Relevant
4. ข้างใต้ทุกลิ้งเว็บไซต์จะมีส่วนที่แสดงรายละเอียดของแต่ละเว็บไซต์โดยจะมี Tab อยู่ 3 ส่วนคือ
   4.1 comment เมื่อผู้ใช้ต้องการแสดงหรือดูความคิดเห็น
   4.2 relevant เมื่อผู้ใช้ต้องการดูอัตราของ relevant และ non-relevant และ Not sure ของแต่ละเว็บไซต์
   4.3 who seen? เมื่อผู้ใช้ต้องการดูว่าเว็บไซต์นี้มีใครเข้าไปดูมาแล้วบ้าง
5. ช่องขวาบนจะแสดงประวัติการเข้าใช้โปรเจกต์ปัจจุบันว่า เข้าเว็บไซต์ไหนอยู่ หรือปิดเว็บไซต์ไหนไปแล้ว รวมถึงการเข้าออกโปรเจกต์ที่ทำงานร่วมกัน
6. ช่องขวาล่างคือ instant message มีไว้สำหรับสนทนากับเพื่อนร่วมทีม

รูปที่ 3 Interface จัดการโปรเจกต์ของระบบ CoZpace

จากรูปที่ 3 แสดงหน้าจัดการโปรเจกต์ของระบบ CoZpace ประกอบด้วย

1. ปุ่ม Create new project ใช้ในการสร้างโปรเจกต์
2. ด้านบนจะมี Tab อยู่ 3 ส่วนคือ
   1.1 Current คือ โปรเจกต์ที่ทำอยู่ในปัจจุบันและยังไม่เสร็จ
   1.2 Completed คือ โปรเจกต์ที่เสร็จแล้ว
   1.3 Invite คือ จะมี คำเตือนขึ้นเตือนให้ กดยอมรับ เมื่อมีคำเชิญจากผู้ร่วมทีมคนอื่นๆ เพื่อเข้าร่วมในโปรเจกต์
3. กล่องด้านขวาจะแสดงรายละเอียดของแต่ละโปรเจกต์ ซึ่งจะมี 2 ปุ่ม (ซ้ายบน) ซึ่งปุ่มแรกจะใช้ในการจัดการเชิญ เพื่อนเข้าโปรเจกต์และปุ่มที่สองใช้ในการเข้าสู่โปรเจกต์นั้น



รูปที่ 4 การแสดงเว็บเพจของระบบ CoZpace

จากรูปที่ 4 Interface แสดงผลเว็บไซต์ที่เปิดปัจจุบัน มีส่วนที่สำคัญคือ

1. 4 ปุ่ม ด้านขวาบน มีไว้จัดการกับเว็บไซต์ปัจจุบันที่เปิดอยู่ โดยจะมีหน้าที่ดังนี้
   1.1 comment ใช้แสดงความคิดเห็นเว็บไซต์ที่เปิดอยู่
   1.2 relevant ใช้ในการให้ rate เว็บไซต์นั้นโดยมี 3 ตัวเลือกคือ relevant, non-relevant และ not sure
   1.3 seen แสดงผู้ร่วมทีมที่เคยเข้ามาดูเว็บไซต์นี้มาก่อน
   1.4 Snapboard ใช้ในการตัด ข้อมูลที่สนใจบางส่วนในเว็บไซต์ปัจจุบัน

รูปที่ 5 ฟังก์ชัน Snapboard ของระบบ CoZpace

จากรูปที่ 5 ฟังก์ชัน Snapboard (กระดานแสดงและแชร์ข้อมูลระหว่างกลุ่ม) เป็นฟังก์ชันใหม่ที่เรานำเสนอขึ้น เพื่อช่วยในการสนับสนุนการค้นหาแบบกลุ่ม ซึ่งสนับสนุนทั้ง synchronous และ asynchronous collaboration ทำให้ผู้ใช้สามารถร่วมมือกันได้ทั้งในช่วงเวลาเดียวกันหรือ ช่วงเวลาที่ต่างกันได้ ซึ่งประกอบด้วยความสามารถที่สำคัญคือ

1. สามารถแชร์รูปให้กับผู้ร่วมค้นหาคนอื่นๆได้
2. การตัดส่วนสำคัญในหน้าเว็บ
3. มีการอ้างอิงถึงแหล่งที่มาของเว็บ
4. สามารถแสดงความคิดเห็นกับรูปได้

5.2 ผลการประเมินคุณภาพและความพึงพอใจ

จากการรวบรวมข้อมูลจากแบบสอบถาม พบว่าผู้ใช้ส่วนใหญ่เห็นว่าระบบ CoZpace ทำให้การค้นหาข้อมูลได้รวดเร็วขึ้น เนื่องจากมีการแสดงเว็บไซต์ที่ผู้ใช้คนอื่นเข้าไปแล้ว รวมถึงการให้ relevant หรือ non-relevant กับเว็บนั้น ก็ทำให้ผู้ใช้ทราบได้ว่าเว็บไซต์นั้นเป็นเว็บไซต์ที่น่าจะเกี่ยวข้องกับเรื่องที่ค้นหา รวมถึงสามารถดูได้ว่าเว็บไซต์นั้นมีผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คนใดเคยเข้าดูแล้วบ้าง ในส่วนของ instant message ผู้ใช้เห็นว่ามีประโยชน์น้อย เนื่องจากระบบนั้นสามารถ comment ในแต่ละเว็บไซต์ได้ ทำให้ผู้ใช้ไม่จำเป็นต้องใช้ instant message มากนัก ส่วนการแสดง keyword ที่เคยใช้ค้นหา ผู้ใช้เห็นว่ามีประโยชน์เพราะช่วยลดการใช้ keyword ที่ซ้ำกับผู้ใช้คนอื่นและช่วยให้ผู้ใช้สามารถดูประวัติ keyword ที่เคยใช้ได้

ในส่วนของ Snapboard ทำผู้ใช้มีความสะดวกมากขึ้นเวลากลับมาดูอีกครั้ง เนื่องจากไม่ต้องเสียเวลาเข้าแต่ละเว็บเพจเพื่อหาข้อมูลในส่วนสำคัญของเว็บไซต์นั้น รวมถึงทำผู้ใช้คนอื่นสามารถดูเฉพาะข้อมูลส่วนที่สำคัญในเว็บไซต์นั้นได้ และข้อเสนอแนะจากผู้ใช้พบว่า มีผู้ใช้บางคนเสนอแนะให้ระบบสามารถสื่อสารกันได้ด้วยเสียง ซึ่งน่าจะช่วยเพิ่มประสิทธิภาพได้และสะดวกในการค้นหามากขึ้น

6. บทสรุป

ในบทความนี้ นำเสนอระบบค้นหาข้อมูลแบบร่วมมือชื่อ CoZpace ซึ่งออกแบบมาเพื่อสนับสนุน การรับรู้ในการค้นหาข้อมูลแบบกลุ่ม ช่วยเพิ่มประสิทธิภาพ ในการค้นหาข้อมูล โดยจะมีฟังก์ชันที่ช่วยให้แต่ละคนสามารถ ร่วมมือกันได้อย่างมีประสิทธิภาพ เช่น การ chat กับผู้ร่วมค้นหา การ comment เว็บไซต์ที่น่าสนใจ การ relevant เว็บไซต์ที่เกี่ยวข้อง และการมี awareness ในระบบ ซึ่งช่วยให้ผู้ใช้รับรู้ถึงกิจกรรมต่างๆ ที่ผู้ร่วมค้นหาได้กระทำไม่ว่าจะเป็น การเข้าเว็บไซต์อะไร หรือใช้คีย์เวิร์ดอะไร ในการค้นหาข้อมูลในปัจจุบัน จะทำให้ไม่เสียเวลาในการค้นหาที่ซ้ำซ้อนกันกับผู้ร่วมค้นหา เช่น การใช้คีย์เวิร์ดเดียวกัน ซึ่งจะได้ผลการค้นหาที่เหมือนกัน หรือการเข้าเว็บไซต์เดียวกันซึ่งอาจเป็นเว็บไซต์ประเภท non-relevant ซึ่งไม่มีประโยชน์และเสียเวลา อีกทั้งระบบมีการเก็บสรุปกิจกรรมที่เราทำในระบบ ซึ่งสามารถดูย้อนหลังได้ว่าเราเคยทำอะไร เช่น ใช้คีย์เวิร์ดอะไรมาแล้วบ้าง เคยเข้าเว็บไซต์ไหนมาแล้วบ้าง หรือมีเว็บไซต์ไหนที่เรา relevant และ non-relevant ไว้บ้าง

เอกสารอ้างอิง

[1] *Wikipedia. Collaborative Search Engine.* [online] May 2013 [cited 2013 Jun 20] Available from: http://en.wikipedia.org/wiki/Collaborative_search_engine

[2] J. Ashley; *Synchronous and Asynchronous Communication Tools*; [online] Dec 2003 [cited 2013 Jun 20] Available from: http://www.asaecenter.org/Resources/articledetail.cfm?itemnumber=13572

[3] Z. Yue, S. Han, and D. He; *A Comparison of Action Transitions in Individual and Collaborative Exploratory Web Search*; In Proceedings of the 8th Asia Information Retrieval Societies Conference, AIRS 2012; Tianjin, China; pages 56-63

[4] M. Bron, J.V. Gorp, F. Nack, L.B. Baltussen, and M.D. Rijke; *Aggregated Search Interface Preferences in Multi-Session Search Tasks*; In Proceedings of the 36th Annual International ACM Special Interest Group on Information Retrieval Conference, SIGIR 2013; Dublin, Ireland; pages 123-132

[5] M.R. Morris; *Collaborative Search Revisited*; In Proceedings of the 16th Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2013; San Antonio, USA.; pages 1181-1192

[6] C. Shah and G. Marchionini; *Awareness in Collaborative Information Seeking*; Journal of the American Society for Information Science and Technology 2010; USA.; pages 1970-1986

[7] R. Capra, A.T. Chen, K. Hawthorne, and J. Arguello; *ResultsSpace: An Experimental Collaborative Search Environment*; Journal of the American Society for Information Science and Technology 2011; USA.; pages 1-4

[8] M.R. Morris and E. Horvitz; *SearchTogether: An Interface for Collaborative Web Search*; In Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST 2007; pages 3-12

[9] J.R. Perez, T. Leelanupab, and J.M. Jose. *CoFox: A Synchronous Collaborative Browser*; In Proceedings of the 8th Asia Information Retrieval Societies Conference; AIRS 2012; Tianjin, China. pages 262-274

[10] B.P. Hogan, *HTML5 & CSS3*; Reading: Addison-Pragmatic Programmers, LLC. 2010

[11] P. Teixeira. *Professional Node.js*; Reading: Addison-John Wiley & Sons, Inc. 2013

[12] *BingDev API Basics*; [online] [cited 2013 Nov 29] Available from: http://www.bing.com/developers/s/apibasics.html

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

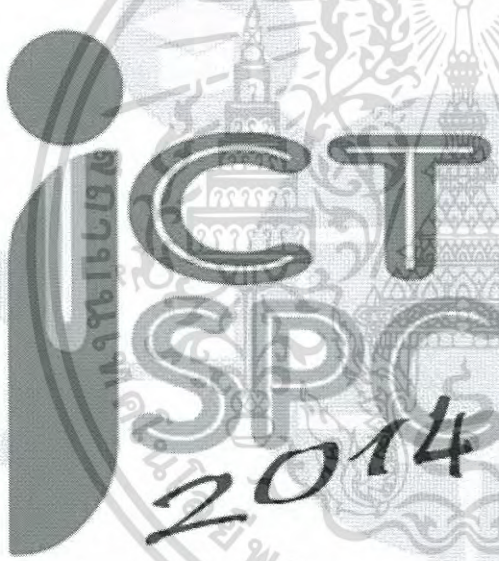
IEEE

2014 Third ICT International Student Project Conference (ICT-ISPC 2014)

March 26-27, 2014

The Faculty of ICT, Mahidol University
Nakhon Pathom THAILAND

Theme : ICT for Better Living



Editor :
Jarernsri L.Mitrpanont

IEEE Catalog Number : CFP14AWE-PRT
ISBN : 978-1-4799-5572-5

TAT UTM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The 2014 Third ICT International Student Project Conference (ICT-ISPC2014)

CoZpace: A Proposal for Collaborative Web Search for Sharing Search Records and Interactions

Hannarin Kruajirayu, Ake Tangsomboon and Teerapong Leelanupab
Faculty of Information Technology, King Mongkuts Institute of Technology Ladkrabang (KMITL)
Bangkok, Thailand. 10520
Emails: hannarin_k@hotmail.com, ake@mercigod.com, teerapong@it.kmitl.ac.th

***Abstract*—Today's search systems are mainly designed for one user to use individually. Search activities in practice can, however, be conducted by two or more users working together. This is due to the fact that information seeking tasks are complex and often requires multiple users' efforts to collaboratively assess and search for relevant information. In some tasks, there is a large amount of information to search or such information requires human intelligence to judge whether it is relevant to information needs. This paper introduces the development of collaborative search system, named "CoZpace". CoZpace is a Web-based application, which allows a pair or a group of users to create a collaborative search task. Within the created task, a user can invite and communicate with other members in a group, share search history records, and mark search results considered relevant. We also propose a newly developed feature, called "snapboard", to support collaborative search activities. With the snapboard, users are allowed to take a snapshot of a part of focused information in a Web site and then share it among collaborators on a display board. A user can mark it as relevant or comment it for further reviews of other collaborators.**

Keywords*—*CoZpace: Collaborative Search Activities; Information Retrieval; Explicit Collaboration; Synchronous and Asynchronous Collaboration

I. Introduction

Searching information on the Web is a non-trivial task. The amount of Web information is considerably large and it is not originally structured for search. Although many commercial search engines are available, they are typically designed for one user to use independently. In fact, information seeking is considered *not* a solitary activity [1]. Two or more users can carry out search by working together. This is because information seeking tasks are complex and require users to review a number of different documents. Users have to explore and assess all aspects of possible relevant information and finalize the outcome of their search into a summary of relevance to the tasks [2], [3]. In these tasks, multiple users can collaboratively work on searching information towards the same goal.

In this paper, we present a new collaborative search system named "*CoZpace*". It is designed for a pair or a group of people with the same information seeking goal to perform search together. Also, CoZpace provides support for group awareness[1]. We also introduce five main features of CoZpace: instant-messaging, commenting on a web page, recording activity history, marking a relevant web site, and taking a snapshot of focused information. This last is a new feature called "*Snapboard*". The system is a Web-base application using HTML, CSS and JavaScript.

The rest of this paper is structured as follows. In section II, we describe background and related work, followed by a survey of related collaborative search systems. Section IV explains a use case diagram and technologies behind the application development. Section V illustrates the architecture and design of CoZpace as well as all its features. We summarize our proposal in section VI.

II. Background And Related Work

A. Dimensions of collaborative search

In collaborative search, the points of view of two dimensions, i.e., time and user's intent, are described as follows.

1) *Time-Based Collaboration*

Synchronous collaboration is a real-time collaboration. Collaborators can work at the same time either in the same place or in the different places. Its immediacy helps users make a plan, solve a problem and make a decision together faster. A possible synchronous communication is, for example, a chat session.

Asynchronous collaboration collects activities of collaborators, e.g., chat history and query history, and then display them to other collaborators who come to search at different times.

2) *User's Intent*

Implicit collaboration is a scenario in which one user can work with others without awareness of collaboration. For this type of collaboration, a search system exploits collaborators' behaviors, e.g., relevant feedback, to improves search results of another user.

Explicit collaboration is a collaboration that users work together directly to carry out search tasks. For example, one user discusses about search plan with other users by instant messaging.

B. Studies in Collaborative Search Activities

Different collaboration styles may greatly affect user's search actions. *Yue et al.* [4] compared the processes of searching activities in three different conditions given as follows: *i*) Collaborative search with explicit communication, *ii*) Collaborative search without explicit communication and *iii*) Individual search.

Despite the increasing of specifically designed tools to support the collaborative web search (e.g., Coagmento [1]

[1]Group awareness is to provide information to all group members about the status of search tasks and the activities of other collaborators at a given time. This does not matter how people search together synchronously or asynchronously.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

and SearchTogether [5]), most users commented that they hardly utilized advanced functions given by them such as video conferencing or shared screen projection. Morris [6] described that even though those functions offer rich collaborative experiences, users prefer using simpler functions for collaboration, e.g., e-mail, texting, instant-messaging, phone calls, and social networking.

III. Related Systems

1) Coagmento: [1] is a Firefox plugin that helps multiple users work in collaboration. It includes a toolbar and a sidebar that allows users to communicate, share and organize information among collaborators.

2) ResultsSpace: [7] is designed for a small group to collaborate asynchronously. Its user interface supports the awareness of previous and ongoing actions of collaborators. *Capra et al.* [7] presented the main features of ResultSpace, including collaborative ratings and displays embedded into search results, a results filtering mechanism, a peripheral display of past queries of collaborators.

3) Search Together: [5] is a prototype that allows remote users to synchronously or asynchronously collaborate while searching information. The system supports users to collaborate both the processes (e.g., formulating queries and choosing result to explore) and products (e.g., commenting on, rating found items and creating a shared summary).

4) CoFox: [8] provides a synchronous collaboration of a pair of users, i.e., the local user and remote user, as if they are co-located. The system has two interfaces: *i*) The local user area allows the local user to perform search by using, for example, tab browsing, search for text on a given page, etc. *ii*) The remote user area is a real time video streaming that shows what the other user is doing.

IV. System Analysis and Development

A. Use Case Diagram

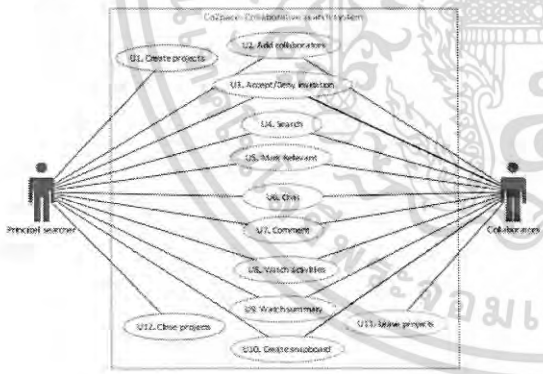


Figure. 1: Use case diagram of CoZpace

We show a use case diagram of CoZpace in Fig. 1. representing the user's interactions with the system and functional requirements. The details of its components are described as follows:

- **Principle Searcher** - This user initializes a project, called a collaborative search task, and can invite any user to join it.
- **Collaborators** - These users share a common information need and wish to join the project. They all work together toward the same goal.
- **U1: Create projects** - This function is for creating the project or collaborative search task.
- **U2: Add collaborator** - A principle searcher or collaboration partners can invite others to join the project.
- **U3: Accept/deny invitation** - After receiving an invitation, a collaborator can accept to join the project or deny it.
- **U4: Search** - Users can perform search by using their natural language in text queries.
- **U5: Mark Relevant** - After finding relevant search results or websites, users can mark them as relevant to the search task/project.
- **U6: Chat** - Users can communicate with everyone in the group by instant messaging.
- **U7: Comment** - This function allows users to comment on a web page.
- **U8: Watch activities** - A user can browse the project timeline that shows all collaborators' previous activities.
- **U9: Watch summary** - A user can see a summary of all activities (i.e., query history, seen websites, snapboard and relevant websites). The summary can be filtered by types of activities.
- **U10: Create snapboard** - Any user can make a snapshot of focused information. The snapshot will then be posted on a display board for sharing among collaborators. Each snapshot is a link that can be clicked to navigate the source of information.
- **U11: Leave projects** - Any collaborator, excluding the principle searcher, can leave the project.
- **U12: Close projects** - Only the principle searcher can close a project when it completed.

B. Technologies

- **Web sockets** [9] defines a full-duplex single socket connection over which messages can be sent between client and server. The Web Socket standard simplifies much of the complexity around bi-directional web communication and connection management.
- **NodeJs** [10] is server-side javascript. It uses the concept of asynchronous event-driven model to support the concurrency in order to reduce the overhead between threads. It is suitable for system to work in "real time".

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Figure. 2: Interface for showing search results

- **Bing API** is a search service that enables developers to embed a search mechanism in their websites or applications. It offers a search functionality for seeking information in multiple source types, e.g., web, images, news and videos. We employ the Bing API to provide web search in the CoZpace.

V. System Design And Implementation

A. Architecture

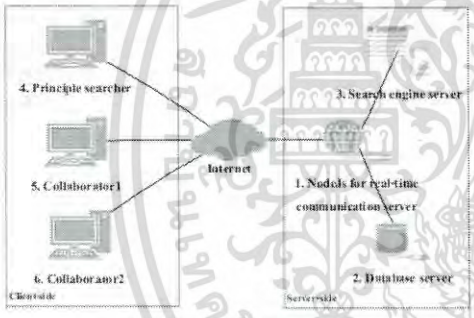


Figure. 3: Architecture of CoZpace system

An architecture of CoZpace and its components are demonstrated in Fig. 3. The details of five components are given as follows:

1) **NodeJs server** - A server that provides services of real-time communication, such as instant messaging and real-time activity stream.
2) **Database server** - A server that collects data of user activities, such as query history, chat history and web history.
3) **Search engine server** - A server that collects web pages to show search results from user's search queries.
4) **Principle searcher** - A user that can create and close a project as well as invite other collaborators into the project to search information together.
5) **Collaborator** - A user that accepted the invitation from principle searcher and afterwards is eligible to invite other collaborators to join the project.

B. CoZpace's design and features

In this section, we illustrate the interfaces of our collaborative search system. CoZpace provides a project timeline that shows history of all collaborators' activities. Moreover, for direct communication, the user can use instant message. The system interfaces are composed of three main components: *i*) Show search results. *ii*) Manage project. *iii*) Show websites

Fig. 2 demonstrates the interface that is designed to show search results. It consists of six sections as follows:

1) Four buttons on the top-left show options to groups of search activities done by everyone in a current project.
   a) **"Seen" button** presents a list of websites that have already been seen by any collaborators and are sorted chronologically.
   b) **"Queries" button** shows the queries that was used by the collaboration partners.
   c) **"Snapboard" button** demonstrates the pictures of anchor links that were snapped by anyone in a group.
   d) **"Relevant" button** displays the searched websites that are judged or marked by group members, i.e., relevant, irrelevant or not sure.
2) The search box allows users to enter a search query, submitted to the Bing API.
3) This section shows search results responding to a user's query. It contains two buttons on the top-right, i.e., thumb up button as "relevant button" and thumb down button as "irrelevant button". These buttons allow the user to mark the website relevant or irrelevant. We use three colors to represent relevant status of documents:
   a) **Green** when the relevant button is pressed.
   b) **Red** when a user presses the irrelevant button.
   c) **Yellow** when a user presses both two buttons, it means not sure.
4) This section presents three tabs that show the details of the searched websites:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

a) **Comment tab** is to receive comments of searched website from collaborators.
b) **Relevant tab** displays a percent of relevant, irrelevant and not sure.
c) **Who's seen? tab** shows a list of collaborators who have already seen the searched website.

5) This section presents the project timeline that shows all collaborators' history, e.g., query that is used, the websited that is opened or closed and the project that is closed.
6) This section provides an instant messaging for the collaborators to talk with one another and exchange their opinions.

Figure. 4: Interface for managing a project

Fig. 4 shows the details of interface for managing project, composed of three sections. These include:

1) Three tabs on the top-left are given as follows,
   a) **Current tab** presents the current projects.
   b) **Completed tab** presents the closed projects.
   c) **Invite tab** shows notifications of the invitation from the principle searcher.
2) Create new project button can be used to create a new project.
3) The box on the right side presents the detail of a current project. Any user can invite anyone to his current projects by clicking on the first button on the top-right and also can access to his current project by clicking the second button.

Figure. 5: Interface for showing websites

We show a CoZpace interface for showing websites in Fig. 5. There are four buttons on the top-right, consisting of:

1) **Comment button** can be used to provide feedback or exchange opinion of a website.
2) **Relevant button** can be used to mark the website relevant or irrelevant.
3) **Seen button** presents a list of the collaborators that have seen the website.
4) **Snapboard button** can be used to take a snapshot in some part of the website.

In Fig. 6, we illustrate the snapboard function of CoZpace. Snapboard supports both synchronous and asynchronous collaborations to help users collaborate both at the same time and at different times. The core of snapboard function is listed as follows:

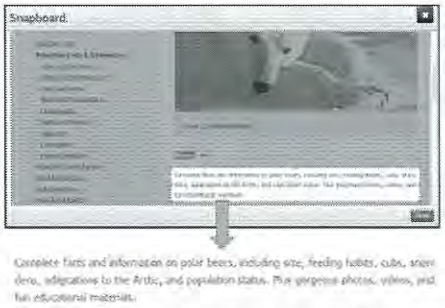


Figure. 6: Snapboard function of CoZpace

1) A user can take a snapshot of the main idea in web page.
2) A user can share a snapshot among collaborators.
3) A user is able to comment in a snapshot.
4) The web address is included in a snapshot.

VI. Conclusion

In this paper, we introduced the CoZpace that was developed by using a Web Socket technology for real-time communication between clients and a server. By doing so, our system accommodates multiple users for "synchronous collaboration" via, for example, instant messaging and live-update of user activities. All user interactions with the system are also recorded in database and showed in a search history. As a result, collaborators can carry out search tasks at different times, enabling "asynchronous collaboration" and increasing awareness of search collaborators. For the future work, we plan to conduct a user study to evaluate the effectiveness of CoZpace.

References

[1] C. Shah and G. Marchionini, "Awareness in collaborative information seeking," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 10, pp. 1970–1986, 2010.

[2] G. Marchionini, "Exploratory Search: From Finding to Understanding," *Communications of the ACM*, vol. 49, no. 4, pp. 41–46, 2006.

[3] R. W. White and R. A. Roth, "Exploratory Search: Beyond the Query-Response Paradigm," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 1, pp. 1–98, 2009.

[4] Z. Yue, S. Han, and D. He, "A comparison of action transitions in individual and collaborative exploratory web search," in *Information Retrieval Technology – the 8th Asia Information Retrieval Societies Conference*, ser. AIRS '12, Tianjin, China, 2012.

[5] M. R. Morris and E. Horvitz, "Searchtogether: An interface for collaborative web search," in *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '07. New York, NY, USA: ACM, 2007, pp. 3–12.

[6] M. R. Morris, "Collaborative search revisited," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW '13, Texas, USA., 2013, pp. 1181–1192.

[7] R. Capra, J. Arguello, A. Chen, K. Hawthorne, G. Marchionini, and L. Shaw, "The resultsspace collaborative search environment," in *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '12. New York, USA, 2012, pp. 435–436.

[8] J. Rodriguez Perez, T. Leelanupab, and J. M. Jose, "Cofox: A synchronous collaborative browser," in *Information Retrieval Technology – the 8th Asia Information Retrieval Societies Conference*, ser. AIRS '12, Tianjin, China, 2012.

[9] B. P. Hogan, *HTML5 and CSS3: Develop with Tomorrow's Standards Today*, 1st ed. Pragmatic Bookshelf, 2011.

[10] P. Teixeira, *Professional Node. js: Building Javascript Based Scalable Software*. Wiley. com, 2012.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Program

THE TWENTIETH INTERNATIONAL SYMPOSIUM ON ARTIFICIAL LIFE AND ROBOTICS

(AROB 20th 2015)

AROB 20th ANNIVERSARY

January 21– 23, 2015

B-Con Plaza, Beppu, Oita, JAPAN

International Society of Artificial Life and Robotics

助成　日本万国博覧会記念基金
Supported by the Japan World Exposition 1970 Commemorative Fund.
この助成金は、日本万国博覧会の収益を基にしています。
公益財団法人　関西・大阪21世紀協会

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

Extracting Visual Snippets for Query Suggestion in Collaborative Web Search

Hannarin Kruajirayu, Teerapong Leelanupab

Knowledge Management and Knowledge Engineering Laboratory
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang (KMITL)
Bangkok, Thailand. 10520

hannarin_k@hotmail.com, teerapong@it.kmitl.ac.th

Abstract: This paper proposes an approach to generate query suggestions by employing information from user-created visual snippets. In order to generate query suggestions, we apply the optical character recognition (OCR) technique to extract a set of words presented in the visual snippet. The natural language processing (NLP) is used to identify the words that might be relevant to users' information need. The goal of this paper is to compare the accuracy of terms extracted by our approach with that by a Lemur search toolkit. In our experiment, we use 1,000 English web pages from a dataset in category B of ClueWeb09. The visual snippets are created from the web pages. Then, their important terms are extracted by our approach. Outcomes of extracting terms for query suggestions are evaluated by using Precision, Recall, and F-measure. The experimental results indicate that our approach effectively identifies and generates useful terms for query suggestion.

Keywords: Query Suggestion, Visual Snippet, ClueWeb09, Lemur Search Toolkit.

1 INTRODUCTION

The visual summarization of web pages has become a useful feature for searching information from the web. Typically, to achieve search and re-finding tasks, users might rely on common features representing previously visited web pages. These web pages are often represented in the forms of page titles in a user's history, a bookmark of URLs, and thumbnails of some part of web pages. These different representations have their own advantages for each task. For example, in different contexts users may require to visit the same web page with different information needs. Therefore, users cannot recognize the thumbnails that are previously captured from some parts of visited web pages. As a result, one form of the common page representations may not be suitable for other tasks. However, the visual snippet is proposed and designed to bring together benefits of typical representations. It combines both textual snippets and thumbnails for a better recognition.

We presented a new feature, "Snapboard", in our collaborative web search system [1], called "CoZpace." Figure 1 demonstrates the Snapboard that is a shared space in the form of a display board that allows a user to share "Visual Snippets" among collaborators. The visual snippet is a snapshot of focused and salient information, captured by the user or collaborators. It acts as a concise and quick human-generated summary, which is potentially relevant to the users' information needs. These users are those who have the same interest and collaboratively search for particular information. The visual snippet in CoZpace is not only useful for emphasizing important information for team collaborations, but also valuable for search and revisitation tasks. By means of visual snippets, users can easily navigate to the same web page having previously viewed, or read information on them and pose new queries based on given information. To support the finding of new information, we propose the use of visual representations of web pages that call out relevant texts for generating query suggestions. This can be regarded as implicit collaboration, since one user unknowingly employs information generated from other users' efforts in an automatic way.

Fig. 1. Example Visual Snippets of CoZpace.

The contributions of this paper are structured as follows. We first describe a brief overview of related works and an

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015)
B-Con Plaza, Beppu, Japan, January 21-23, 2015

approach to generate query suggestions in the next section. Section 4 explains experiment design to evaluate the performance of our approach, followed by the results discussion. We then describe conclusion of our evaluation in section 6.

2 RELATED WORK

In 2009, Teevan et al. [2] presented visual snippet that is a summarization of web pages for search and re-finding tasks. In the experiment, they compared the performance of visual and textual snippet in three main types of search tasks: homepage finding, shopping and medical tasks. Their outcomes demonstrated the visual and textual snippet not much of a difference in speed for searching new web pages, which have never been seen by the users, and the visual snippet is faster than textual snippet for re-finding the previously seen web pages. In 2010, Jiao et al. [3] continued to explore visual snippet of web pages. They described about generating the internal image for visual snippet, which is a representative of the content in the web pages. It is difficult work to create an internal image when the internal image of a web page is unavailable or the web page has a complex layout and rich content. They proposed a new strategy, which retrieves the representative image from the external entire Internet to summarize web page, called external images. Moreover, their study investigated not only how different internal and external image summarizations perform in search and revisitation, but also how different internal and external image summarizations perform for different kinds of web pages. Their results demonstrated that the internal images could provide a reliable summarization on web pages with dominant images and web pages with simple structure and the external images were a useful information to complement the internal images and very useful in helping users understanding new web pages. However, in 2010, Aula et al. [4] compared the efficacy and effectiveness of visual and textual information summaries for web page previews. Their results suggested that the visual summarizations perform worse than the textual summarization. Moreover, they found that the textual summarizations tended to make the users systematically overestimate the helpfulness of web pages. The best performance was obtained by combining textual and visual information summarizations. The combination by placing title and URL below the visual summarization makes users focus on both visual and textual information.

Optical Character Recognition (OCR) [5] handles a problem of recognizing optically processed characters. OCR is the conversion of images of typewritten or printed text into machine-encoded/computer-readable text. The performance of OCR is directly dependent upon the quality of the input image. Recently, OCR tools became commercially available in the markets, but only few of them are open source. The Tesseract [6] is one of the popular open sources OCR software [7]. It was developed by HP Labs Bristol in between 1984 to 1994. The Tesseract was modified with greater accuracy and concentrated more on improving rejection efficiency than on base-level accuracy.

Most existing search engines are able to automatically suggest a list of textual query terms based on users' current query input called Textual Query Suggestion. In 2009, Zha et al. [8] proposed a new query suggestion scheme named Visual Query Suggestion (VQS) which was dedicated to image search. They demonstrated that VQS was able to more precisely and more quickly help users specify and deliver their search intents. Denney et al. [9] (2009) introduced a concept of related search suggestion for searching video. When a user puts a query for video searches, their system not only displays the video search results but also shows related video search of the query. In 2012, Bian et al. [10] also studied Visual Query Suggestion. They proposed a new query suggestion scheme named Visual Query Attributes Suggestion (VQAS) for image search. VQAS suggests a list of informative attributes based on users' query image. These attributes reflect the visual properties and key components of the query, which is able to help users specify and deliver their search intention in a more precise way.

3 APPROACH

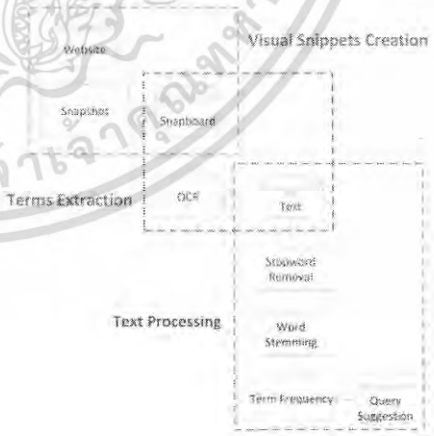


Fig. 2. The framework of our approach.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

As shown in Figure 2, our approach is composed of three main steps: i) Visual snippets creation, ii) Terms extraction, and iii) Text processing.

3.1 Visual Snippets Creation

The first stage of visual snippets creation is to retrieve the English web pages. Then, we generate quality visual snippets in PNG image format by using an open source Python tool called Webkit2png [11].

3.2 Terms Extraction

In order to get more exploits on visual snippet, we explore important terms in the visual snippet by employing the OCR tool called Tesseract to convert text image in visual snippet into computer-readable text. However, it is usually difficult to extract text from visual snippet because of the different size, different style, and complex background of visual snippet. Therefore, the effectiveness of OCR depends on image pre-processing. We adjust the quality of visual snippet by increasing image contrast, increasing image size, and changing image color to black & white, as an example in Figure 3.

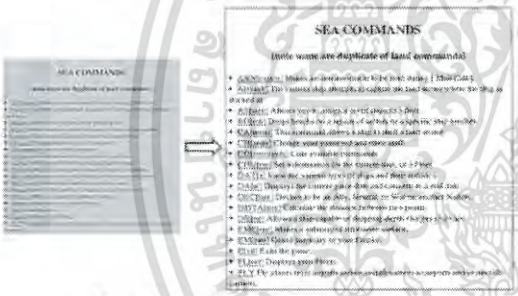


Fig. 3. Example adapting visual snippet.

3.3 Text Processing

First of all, frequently occuring and insignificant words, called stopwords, e.g., a, an, and the, are removed from the extracted text. Then, we remove inflectional and derivational suffixes in order to combine word variants into the same stem or root, called words stemming, e.g., processes, processed, and process can be shorten to process. Finally, we also use the term frequency (TF) as the number of times that a term occurs in visual snippet. If a term has high TF, the term might be relevant to users' information need.

4 EXPERIMENT DESIGN

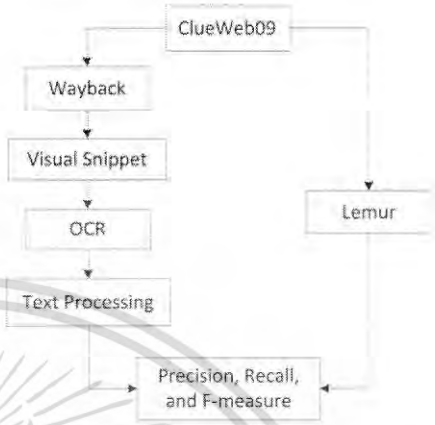


Fig. 4. The framwork of experiment design.

We conduct an experiment to compare the performance of our approach with that of Lemur search toolkit. In our experiment, we only focus on the English Web pages from ClueWeb09 dataset. We use Wayback tool to access ClueWeb09 and Webkit2png to create a visual snippet. We then employ OCR technique to recognize terms in the visual snippet followed by text processing. In order to assess effectiveness of our approach, we use the most prevailing measures that are precision, recall, and F-measure, as shown in Figure 4.

4.1 ClueWeb09 Dataset

The ClueWeb09 [12] dataset was created to support research on information retrieval and related human language technologies. It consists of about 1 billion web pages in ten languages that were collected in January and February 2009. The dataset is used by several tracks of Text REtrieval Conference (TREC) [13] which is a series of workshops focusing on a list of different information retrieval (IR) research areas. As we found out that all dataset in Category B contains only English web pages, We randomly select one dataset in categories B of ClueWeb09 consisting of about 30,000 English web pages. As English web pages in our experiment are stored as the Web ARChive (WARC) file format [14] in ClueWeb09 dataset, the first stage of visual snippets creation is to retrieve the English web pages by using Wayback tool. A simple logical overview of the Wayback Web application in Figure 5 shows that Wayback tool is used to index the WARC file and keep the index in the CDX file [15].

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

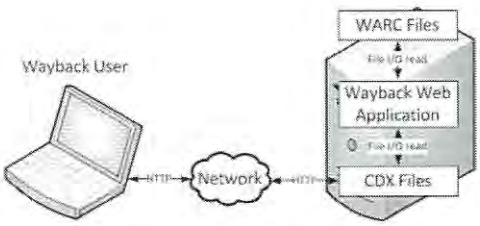


Fig. 5. A logical overview of a Wayback Web application.

To evaluate the overall performance, we use randomly selected 1,000 web pages. However, we found that the number of terms in the majority of web pages was varied from about 100 to 400 terms. Consequently, in order to check whether the number of terms impacts the performance, we also check performance separately from 4 different groups of web pages that are classified by the number of terms: i) 50-200 terms, ii) 201-400 terms, iii) 401-600 terms, and iv) more than 600 terms. Each group contains randomly selected 250 web pages.

4.2 Lemur Search Toolkit

Lemur is a toolkit [16] designed to facilitate research in language modeling and information retrieval. The Lemur toolkit [17] is effective to support the construction of basic text retrieval systems, it comes with all the source code and make files necessary to build the libraries for indexing and retrieval. Accordingly, we index the English web pages by using the Lemur toolkit and take its result as a baseline to compare with the number of terms extracted by our approach. As we assume that the result from the Lemur toolkit is a correct result, precision, recall and F-measure are used as measures for the comparison. Table 1 shows the number of terms and the number unique terms from the indexing by the Lemur toolkit.

Table 1. Statistics of experimental collections.

Name	Docs	Terms	Uniq. Terms
ClueWeb09	1,000	1,307,646	439,227

4.3 Precision, Recall and F-measure

Precision, recall, and F-measure are the basic measures used in evaluating search strategies. We apply precision, recall, and F-measure to evaluate relevant terms extracted from visual snippet. Figure 6 illustrates that we divide the extracted terms to 3 sets for the evaluation as follow:

T1: The set of terms that are extracted by our approach and also indexed by Lemur toolkit.

T2: The set of terms that are indexed by Lemur search toolkit but not by our approach.

T3: The set of terms that are extracted by only our approach.

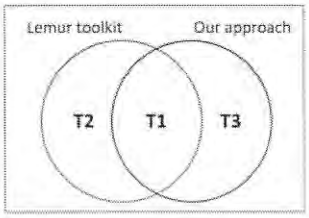


Fig. 6. The set of T1, T2, and T3 for evaluation.

Precision is the ratio of the number of terms that are extracted by our approach and also indexed by Lemur toolkit (T1) to the total number of terms that are extracted by our approach (T1+T3).

$$Precision = \frac{N_{T1}}{N_{T1}+N_{T3}} \quad (1)$$

Recall is the ratio of the number of terms that are extracted by our approach and also indexed by Lemur toolkit (T1) to the total number of terms that are indexed by Lemur toolkit (T1+T2).

$$Recall = \frac{N_{T1}}{N_{T1}+N_{T2}} \quad (2)$$

F-measure is useful in giving a single numeric measure of overall performance combining both precision and recall measure.

$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

5 EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results in Table 2 shows that the terms extracted by our approach match with 81.44% of the terms in baseline (T1). It means that the performance of our approach to retrieve terms is very high similar to Lemur search toolkit. Even though, there are a large number of terms in T3 which is 28.63% of baseline, we cannot judge that they are incorrect. Because, Lemur may not be comprehensive to retrieve all the terms from ClueWeb09. Still, the terms extracted by our approach may possibly be useful for generating query suggestion. Moreover, it can be observed that our approach can achieve the precision, recall

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The Twentieth International Symposium on Artificial Life and Robotics 2015 (AROB 20th 2015),
B-Con Plaza, Beppu, Japan, January 21-23, 2015

and F-measure rate up to 73.9%, 81.4%, and 77.5%, respectively.

Table 2. The result of 1,000 Web pages

T1	357721 terms (81.44%)
T2	81,506 terms (18.56%)
T3	125,765 terms (28.63%)
Precision	0.739
Recall	0.814
F-measure	0.775

The precision, recall and F-measure in Table 3 shows that our approach performs well in every range of the number of terms. Moreover, there is no signal of impact from the different of the number of terms to the performance of our approach because the precision, recall and F-measure in the four groups are only a bit different.

Table 3. Precision, recall and f-measure of the web pages was classified to four group

Group of terms	Precision	Recall	F-measure
50-200 terms/page	0.709	0.836	0.767
201-400 terms/page	0.719	0.837	0.774
401-600 terms/page	0.717	0.828	0.768
>600 terms/page	0.767	0.796	0.781

6 CONCLUSIONS

In this work, we proposed an approach to generate query suggestion by exploiting information in visual snippet. We conducted an experiment to evaluate our approach by using ClueWeb09 dataset, Wayback, Webkit2png, Lemur search toolkit, precision, recall, and F-measure. The experiment results demonstrate that our approach has a high precision, recall, and F-measure. In other words, our approach is capable of generating satisfactory and useful terms for query suggestion, which is relevant to users' search intention.

REFERENCES

[1] H. Kruajirayu, A. Tangsomboon, and T. Leelanupab. Cozpace: A proposal for collaborative web search for sharing search records and interactions. In *The 2014 Third ICT International Student Project Conference*, ICT-ISPC '14, 2014.

[2] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. Andr´e, and C. Hu. Visual snippets: Summarizing web pages for search and revisitation. In *Proc.of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, 2009.

[3] B. Jiao, L. Yang, J. Xu, and F. Wu. Visual summarization of web pages. *Proc.of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 499–506, 2010.

[4] A. Aula, R. M. Khan, Z. Guan, P. Fontes, and P. Hong. A comparison of visual and textual page previews in judging the helpfulness of web pages. In *Proc.of the 19th International Conference*, WWW '10, 2010.

[5] S. Mori, H. Nishida, and H. Yamada. *Optical character recognition*. John Wiley & Sons, Inc., 1999.

[6] GOOGLE. Tesseract-ocr. http://code.google.com/p/tesseract- ocr/, 2012.

[7] R. Smith. An overview of the tesseract ocr engine. In *Proc.of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, ICDAR '07, 2007.

[8] Z. J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang. Visual query suggestion. In *Proc.of the 17th ACM International Conference on Multimedia*, MM'09, pages 15–24, 2009.

[9] J. Dennoy, J. Hamilton, T. C. Hoad, K. R. Mcdonald, and R. J. Qian. Visual query suggestions, June 27 2008. US Patent App. 12/163,655.

[10] J. Bian, Z. J. Zha, H. Zhang, Q. Tian, and T.-S. Chua. Visual query attributes suggestion. In *Proc.of the 20th ACM International Conference on Multimedia*, MM '12, 2012.

[11] P. Hammond. Webkit2png. www.paulhammond.org/webkit2png/, 2013.

[12] J. Callan, M. Hoy, C. Yoo, and L. Zhao. Clueweb09 data set, 2009.

[13] D. Harman. Overview of the first trec conference. In *Proc.of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, 1993.

[14] I. ISO. 28500: 2009 information and documentation-warc file format. *International Organization for Standardization*, 2009.

[15] B. Tofel. Wayback for accessing web archives. In *Proc.of the 7th International Web Archiving Workshop*, 2007.

[16] P. Ogilvie and J. P. Callan. Experiments using the lemur toolkit. In *TREC*, volume 10, pages 103–108, 2001.

[17] J. Allan, J. Callan, K. Collins-Thompson, B. Croft, F. Feng, D. Fisher, J. Lafferty, L. Larkey, T. N. Truong, P. Ogilvie, et al. The lemur toolkit for language modeling and information retrieval. 2003.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The 2015 Asian Information Retrieval Societies Conference (AIRS 2015), was the eleventh instalment of the conference series, initiated from the Information Retrieval with Asian Languages (IRAL) workshop series back in 1998 in Korea. The conference was held in 2-4 December, 2015 at the Queensland University of Technology (QUT), Brisbane, Australia.

The annual AIRS conference is the main information retrieval forum for the Asia-Pacific region and aims to bring together academic and industry researchers, along with developers, interested in sharing new ideas and the latest achievements in the broad area of information retrieval. AIRS 2015 enjoyed contribution spanning theory and application of information retrieval, both in text and multimedia.

This year received 92 submissions form all over the world. Submissions were peer reviewed in a double-blind process by at least three international experts, with at least one senior meta-reviewer. The final programme of AIRS 2015 featured 29 full papers (32%) divided in 10 tracks: "Efficiency", "Graphs, Knowledge Bases and Taxonomies", "Recommendation", "Twitter and Social Media", "Web Search", "Text Processing, Understanding and Categorization", "Topics and Models", "Clustering", "Evaluation", and "Social Media and Recommendation". The programme also featured 8 short papers and 3 demonstrations (12%).

AIRS 2015 also featured two keynotes: the first by Peter Bruza (Queensland University of Technology), titled "Quantum Haystacks Revisited", and the second by Peter Bailey (Microsoft Research), titled "The Great Search Bake Off".

The Conference and Program Chairs of AIRS 2015 extend our sincere gratitude to all authors and contributors to this year's conference. We are also grateful to the Program Committee for the large reviewing effort that guaranteed AIRS 2015 could feature a quality program of original and innovative research in information retrieval. Special thanks go to our sponsors for their generosity: the Australian E-Health Research Centre (CSIRO), the Electrical Engineering and Computer Science School and the Information Systems School of QUT, and Springer. We also thank the Special Interest Group in Information Retrieval (SIGIR) for supporting AIRS by granting in-cooperation status and sponsoring the student travel grant that contributed to pay the travel costs for 10 students presenting their research work at AIRS 2015.

Guido Zuccon, Queensland University of Technology, Australia
Shlomo Geva, Queensland University of Technology, Australia
Conference General Chairs

Hideo Joho, University of Tsukuba, Japan
Falk Scholer, RMIT, Australia
Program Chairs for Full Papers

Aixin Sun, Nanyang Technological University, Singapore
Peng Zhang, Tianjin University, China
Program Chairs for Short Papers and Demonstrations

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Snapboard: A Shared Space of Visual Snippets
A Study in Individual and Asynchronous Collaborative Web Search

Teerapong Leelanupab, Hannarin Kruajirayu, Nont Kanungsukkasem

Faculty of Information Technology
King Mongkuts Institute of Technology Ladkrabang (KMITL)
Bangkok, Thailand. 10520
teerapong@it.kmitl.ac.th, hannarin_k@hotmail.com, k.nont@outlook.com

Abstract. People often engage in many search tasks that be collaborative, where two or more individuals work together with the joint information needs. We introduced and built *CoZpace*, a web-based application that enables a group of users to collaborate on searching the web. We also presented the main feature of CoZpace, named *Snapboard*, which is a shared board for a collection of group-created visual snippets. The visual snippet is a snapshot of focused and salient information captured by a user. It acts as a visual summarization of web pages, which allows any user to quickly recognize information and to revisit web pages. This paper describes example usage scenarios and initially investigates the ways Snapboard facilitates users in individual and asynchronous collaborative search. We then analyze users' interactions and discuss how Snapboard supports search collaboration among study participants.

Keywords: Collaborative Web Search, Visual Snippet, Search User Interface, Exploratory Search, Collaborative Information Behavior

1 Introduction

Although technology behind search engines such as Google, Yahoo! Search has advanced over the years, the designs of their web search interfaces have still largely remained for solitary information seeking; they are typically created for one person to use independently. Relatively, little support is available for a group of people, who share the same information need, to collaborate on search tasks. Examples of such tasks include planning a holiday with family members, organizing a social event with friends, or working on an annual report with colleagues. Additionally, some search tasks are too complex or difficult to be completed by a sole person, such as finding possible areas where the MH370 crashed or listing past and ongoing missions of New Horizons space probe to Pluto. These tasks require input and expertise from others. In fact, information seeking is considered collaborative [16]; multiple users can carry out search by working together. Golovchinsky et al. [7] defined four dimensions of the collaboration model, i.e., intent (explicit vs. implicit), depth of mediation (from user-interface only

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

to deeper algorithmic mediation), concurrency (synchronous vs. asynchronous) and location (co-located vs. remote).

We created a web-based application, CoZpace, for collaborative web searching. It enhances users' awareness to understand the activities of others engaging in a cooperative effort through instant messaging, result assessment, activity history and so forth. Users are aware of the degree to success in their search task, benefit from the work done previously or concurrently by others, and can plan subsequent activities for querying the current set of retrieved web pages. Importantly, we also addresses a key investigated feature of CoZpace, namely Snapboard, which helps a group of search partners to identify relevant information and re-finding the right web pages. Snapboard collects all visual snippets captured by users during either asynchronous or synchronous search sessions. By putting the visual snippets all together within the Snapboard, we allow users to review important web contents in one place without losing their search context. Users can use the visual snippets to evaluate the relevance of web pages together if they each cannot conclude the relevance alone. Focused information can be promptly recognized and its source can be easily revisited.

The type of collaboration, supported by CoZpace, is explicit in that the intention of the users is to work together on search, and the mediation is mainly supported by features in the interface but not algorithms[1]. Although CoZpace can also support both asynchronous or synchronous collaboration, for the evaluations that we describe here we focus on asynchronous collaboration among users and intentionally leave other types and dimensions of search such as co-located and synchronous collaboration for future investigation. Also, we examine the individual use of Snapboard done by a single user. Built as a web application, CoZpace supports both co-located and remote search. Even though location is not our interest here, we assume in our study that users remotely collaborate where they are not allowed to talk with each other during the experiment. We conducted a laboratory-based user study to evaluate the supportiveness of *Snapboard* in individual and asynchronous collaborative search.

2 Related works

2.1 "Active" Collaborative Search Systems

Much research in the area of collaborative search focuses on *active* forms of collaboration (e.g., searching together by being aware of search partners' activities, by being able to plan search tactics, and by working as a team to achieve a shared goal.) An example from this line of work includes a very early system, SearchTogether, which allows remote users to synchronously and asynchronously collaborate on search [13]. It keeps a common history of search queries entered by any group members. Each user is provided with her split search result tab and allowed to communicate with other group members via integrated messaging. Similar to SearchTogether, CollabSearch [20] supports both implicit and

[1] Our previous work experimented an algorithm that generates query suggestions extracted from terms present in visual snippets [10].

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

explicit communication. It has an additional workspace where group members can share the saved whole web pages and textual snippets generated by Google API. CoSearch [1] leverages additional available devices such as extra mice and mobile phones for co-located collaborative web search. It assists users in control and division of labor while jointly searching the web by a single computer.

WeSearch [14] is an interactive tabletop display for face-to-face collaboration. It is designed to support co-located collaborative web search, browsing, and sense-making among a group of up to four people. Perez et al. [15] present a CoFox system, which shows a live video stream of a remote user's search screen to a local user to increase awareness of synchronous search collaboration. In multimedia retrieval domain, a grouping interface for video search, ViGOR [8], is built to assist asynchronous collaboration between users. The main feature of ViGOR is the provision of a workspace for creating and organizing groups of related videos. Each group can have multiple annotations and be used as a starting point for further search queries.

2.2 Studies of Search Habits

Also referred to exploratory search [19], *informational* search is identified as a common web search activity that would benefit from collaboration [3]. In this class of searches[2] [5], a user aims to seek some information on one or more web pages. To achieve this, informational search potentially involves multiple refinements of query terms and often spans many search sessions, and those are what CoZpace is designed for. Yue et al. [20] analyzed transitions of user search actions under three different conditions by varying two search factors (i.e., collaborative vs. individual and with vs. without explicit communication.)

2.3 Summarization of Web Pages

Presenting search results as a summarization of each web page is considered useful for information search, where users can quickly judge which ones of these results are of their interest [18]. For re-finding information, users can also use this similar summarization to access previously visited web pages, usually saved in the bookmarks of a web browser [6]. For simplicity and compactness, the summarization of web pages is typically represented as a textual snippet, consisting of its page title, URL and short text summary of web contents. Nevertheless, reading textual summarization is often time-consuming and difficult to comprehend its information if the textual snippet is very short.

As the classic quote states "A picture is worth a thousand words", this is simply because images convey information that words cannot capture. As such, search and re-finding tasks can be facilitated if the web pages and, in particular, focused information are visually summarized since users can get a quick understanding by seeing an image than reading text. Visual summarization has been studied in several research communities [2, 9, 17] for search, revisitation and bookmarking of web pages.

[2] The remaining two classes of searches are *navigational* and *transactional*, where users aim to find a single specific website or to perform some web-mediated activity, respectively.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Fig. 1. Interface of CoZpace.

3 Snapboard in Collaborative Search

3.1 CoZpace: A Collaborative Search System

CoZpace is a web-based application, which allows a group of remote users to create a collaborative search task. Within the created task, a user can invite and communicate with other members in the group. Fig. 1. illustrates the interface of CoZpace that returns search results after a user submitted an example query, "Restaurants in Kansas City". Seven parts of the screen are described as follows:

1. The task summarization tab that includes four clickable image icons for: *i)* relevance awareness, presenting three categories of judged web search results as indicated by any users in a group, i.e., relevant, non-relevant, and not sure[3], *ii)* Task Snapboard collecting all group-captured visual snippets, *iii)* query awareness, showing a history of all used queries, and *iv)* view list, displaying all web search results clicked to view by any users, respectively;
2. The search bar for entering search query, which also suggests alternative queries and remind for the queries[4] that have already been used in the task;
3. The relevant buttons for marking search results considered relevant (thumbs-up) or non-relevant (thumbs-down);
4. *Per-document Snapboard* showing multiple visual snippets captured from a single document/web page;
5. The comment bar for a user to leave comments about a website;
6. The project timeline showing real time stream of activities; and
7. The instant messaging for real-time discussion among the collaborators.

[3] Web pages are marked as "not sure" if there is no consensus on relevance judgments/voting by group members.

[4] Already used queries are highlighted in yellow as part of query suggestion.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

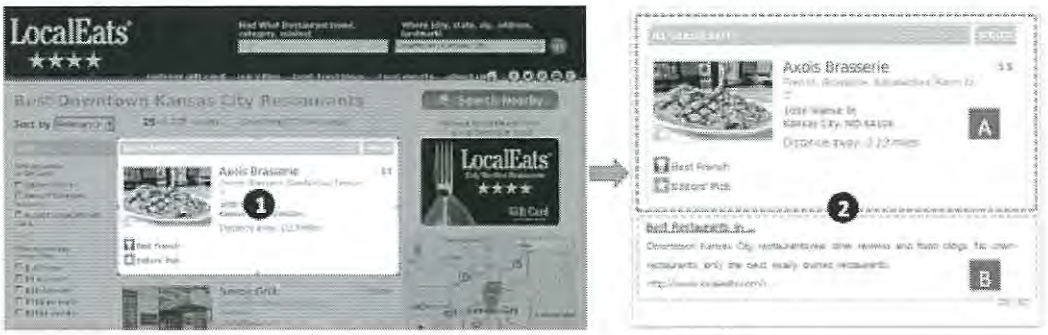


Fig. 2. An example of creating a visual snippet.

Note that all features in CoZapce are interactive and real-time; all interactions of group members are recorded and instantly shown according to their types of summarization. Furthermore, any search results that had been clicked to view by any users in a group will be highlighted in yellow. Our aim of providing the summarizations of different aspects is to facilitate awareness of a shared search task being pursued by other group members, so that unnecessary redundancy of effort can be avoided. For further details about the complete design and features of our CoZpace, we refer interested readers to [11].

3.2 Snapboard

Our Snapboard feature applies visual snippet which is an attractive representation and summarization of a web page, used in Human-Computer Interaction (HCI) to share a relevant part of a web page. However, our visual snippet is different from visual snippets in HCI that are a part of a web page captured by system which might not be relevant to user's information need. Our visual snippet allows a user, in a self-managed manner, to capture a part of focused or salient information in a web page by herself. Moreover, as Aula et al. [2] indicated that a thumbnail which shows the whole web page in a small picture can make users underestimate the relevance of the page and only textual summaries can make users overestimate, it makes us concern that using only a picture in visual snippet might face with the same problem as using thumbnail. Therefore, we allow to include both textual and visual information in our visual snippet.

Fig. 2 displays an example of taking a snapshot of focused information as a visual snippet (1) and its visual snippet (2) generated from a template given a salient snapshot and other metadata of its source/web page, i.e., title, common textual snippet from a search engine[5], and URL of the page. Hovering a mouse over the upper part of the visual snippet (2A) will show a preview of its corresponding snapshot while clicking the lower part (2B) will open the link to the source of the visual snippet in a new web browsing tab within CoZpace.

In CoZpace, visual snippets are shown in a board, called *Snapboard*, which shares them among all members in the same group. There are two types of Snapboard, i.e., for the whole task and per-document. Task Snapboard demonstrates

[5] In this study, we use Bing API for search and textual snippet generation. Other open source search engines can alternatively be used, such as Lemur Indri, Lucene and Terrier toolkits

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Fig. 3. Task Snapboard: a shared space of all visual snippets in a group.

all visual snippets in a single tab (See Fig. 3). Per-document Snapboard displays all visual snippets created from an individual web page (See Fig. 1 (4)), implemented as part of search results where a user can expand or hide it.

4 Experimental Design

In this section, we describe our experimental methodology based on the model of interactive information retrieval evaluation as suggested by Borlund [4]. This user study is conducted to answer the following three research questions:

RQ1: Does our proposed Snapboard feature support exploratory search tasks that require collaboration of two or more users to search together?

RQ2: For an individual user, does the Snapboard also support such search tasks to search alone?

RQ3: Do visual snippets support users in collaboration in a search task when they asynchronously search the web?

4.1 Search Systems

In this experiment, two variations of CoZpace use Bing Search API[6] to provide search and query suggestion functionalities. We restrict the search results in CoZpace by showing only web pages in English. The two systems are: *i)* CoZpace with a Snapboard function (S1) as an experimental condition; and *ii)* CoZpace without the Snapboard (S2) as a control condition.

4.2 Search Tasks

All exploratory search tasks used in this study are selected from the top four search tasks on which people tend to cooperate as surveyed by Morris [12].

(T1) Travel planning: All participants have to imagine that they are planning a trip to the Kansas City, USA. They want to search for information about how they will spend their vacation in USA. Their goal is to find where they will stay, what they will do, and how will they get there, etc.

[6] `http://www.bing.com/dev/`

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(T2) General shopping: All participants have to imagine that they are given USD 30,000 to buy a car. Their goal is to find the technical specifications of cars, brands of cars, and stores which sell cars, etc.

(T3) Literature search: All participants have to imagine that they are assigned to write an article about the US civil war. Their task is to find causes of the civil war, economic causes, consequences of the civil war, civil war effects in the present, and weapons used during civil war, etc.

(T4) Technical information: All participants have to imagine that they want to reduce the use of air conditioner in their house. Their task is to find the best material to use if the purpose is cooling down the roof so that the house temperature remains low, and other solutions regarding roof coating, etc.

4.3 Participants

All participants were 32 volunteers (12 males and 20 females); 12 participants were high school students in an English program from Nakhonnayok Wittayakom school, 10 participants were undergraduate students and the rest were graduate students from the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang (KMITL). All of them are highly proficient in English and web searching. According to the entry questionnaire, they often search more than twice a day. Most of them used to collaborate with other people in web searching by using social networks. Besides, they searched together in a group of three people, on average. However, they have never used a search system that is specifically designed for collaborative search before. All participants are then paired into 16 groups that will asynchronously collaborate on given search tasks. In each pair, we call "P1" a participant who firstly performs search tasks with fresh sessions that do not contain any summarizations and search histories, and call "P2" the other participant who pursues search tasks formerly done by P1.

4.4 Experimental Scenarios

We study the effectiveness of Snapboard in two search scenarios as follows:

(IN) Individual search: In this scenario, a participant (P1) is assigned to search individually. He/She needs to find and mark as many relevant documents as possible, and then gives a reason why the documents are marked as relevant by taking a snapshot (in S1) or commenting (in S2).

(AC) Asynchronous collaborative search: Similar to IN, a participant (P1) in this scenario begin by the first session to perform a search task. Afterwards, the next participant (P2) continues to perform the same search task as P1. Review session starts immediately after finishing each search task.

Review session is a session that participants review the results from a previous session of the same task. It is designed to let everyone use our Snapboard. For review session, in S1, P1 is assigned to review only the visual snippets (Snapboard) of the web pages that marked as relevant by P2 but not marked by P1. In S2, P1 is assigned to review the comments of the web pages instead. Then, P1 has to judge whether the web pages are relevant or not. Afterwards, P2 is also assigned to review and judge the web pages that marked as relevant by P1 but not marked by P2.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 Experimental Procedure

In the experiment, a pair of study participants are asked to collaboratively perform, in asynchronous manner[7], four simulated exploratory search tasks using the two collaborative search systems. To neutralize the effect of human learning behavior in our experiment, we applied a Graeco-Latin Square design to control and rotate the sequence of blocking factors (i.e., systems and search tasks).

In total, our experiment lasted around three hours. The experiment started with an individual introductory session, where participants were given an information sheet, asked to fill in an entry questionnaire, and demonstrated how to use the two search systems. This introduction took approximately five minutes, and was followed by a training session, where each participant was allowed up to ten minutes of interaction and familiarization with the systems. After training, they were asked to perform four exploratory search tasks, as described in Sect. 4.2. For each task, they had a maximum of fifteen minutes to carry it out, followed by a five-minutes review session. After finishing each task, they have to answer a post-task questionnaire. Also, after finishing all tasks, they have to complete an exit questionnaire. After two tasks, a five minute break was given to the subjects, as required by the ethical regulations at KMITL.

4.6 Data Collections

In order to achieve and answer our research questions, we collect both qualitative and quantitative data. Two methods are used to collect the qualitative data, i.e., *i)* open-ended questionnaires, and *ii)* focused group discussions. We collect all logs from participants' interactions and answers from close-ended questionnaires, and then analyze them to be quantitative data, such as, the number of submitted queries, the number of viewed web pages, the number of websites that are marked as relevant, the number of comments, the number of snapshots and the rating scores of user feedback in different aspects.

5 Results and Discussion

5.1 The effectiveness of Snapshot in Individual and Asynchronous collaborative search

Table 1 presents a comparison of the performance between two systems in CoZpace, i.e., S1 and S2, operated in individual search (IN) and asynchronous collaborative search (AC) conditions (Cond). The results show that the average number of visual snippets (Vsnip) made in S1 is obviously higher than that of comments (Comt) in S2 in both IN and AC. It can be interpreted that the participants prefer taking the snapshots as visual snippets rather than commenting to summarize the web pages. The average number of queries used in S1 is higher than that in S2 in both IN and AC. However, it is not necessary to be inferred that participants make more effort to formulate queries in S1 than in S2. The reason behind this outcome is that the feedback from our participants in Table 2 shows that all tasks are easy to formulate query.

[7] Any search session in which one participant performs first is considered the IN scenario and we call such a participant "P1".

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table 1. User interaction statistics – mean (averaged over 16 participants/groups) and standard deviation (in bracket) in Individual and Asynchronous collaborative search for all search tasks.

					Rel Web		Viewed Web		% Uniq Rel
Cond	Sys	Vsnip	Comt	Query	All Rel	Uniq Rel†	All View	Uniq View†	per Uniq View
IN	S1	**20.81** (4.25)	-	**13.69** (3.57)	**12.50** (3.33)	**8.44**	**23.19** (5.15)	14.38	**58.70**
	S2	-	15.88 (5.23)	11.31 (2.72)	11.75 (4.14)	7.88	23.06 (7.99)	**14.44**	54.54
AC	S1	**33.50** (4.37)	-	**22.38** (3.80)	**25.75** (3.61)	**13.88**	**53.50** (6.18)	**26.69**	**52.26**
	S2	-	28.25 (4.64)	20.44 (2.78)	18.69 (3.46)	10.44	49.31 (7.42)	25.31	41.25

† Note that we do not analyze standard deviation of *unique* relevant and viewed web page as they are summed over participants/groups before being averaged.

Furthermore, Table 1 shows that S1 outperforms S2 in term of the average number of web pages marked as relevant (All Rel). In AC, the Rel Web are counted from only the web pages that are marked by both participants P1 and P2 in the same team to be more sure about the relevance of the websites. In addition, the difference between the average number of relevant unique web pages (Uniq Rel) and the All Rel, which is really high in IN and higher in AC. This result shows that many relevant web pages are marked by more than one participants in IN or more than one groups of participants in AC. As a result, we can be more confident that marking the relevant web pages by participants does not happen by chance.

We also check how many percent of all unique web pages clicked to view by participants (Uniq View) are unique relevant web pages (Uniq Rel). The "% Uniq Rel per Uniq View" in S1 is higher than that in S2 in both IN and AC, indicating that S1 is more effective than S2 in the support of users to find relevant web pages from the returned and clicked web pages. From the above results, we can answer RQ2 and RQ3 that the Snapboard support users in both individual and asynchronous collaborative search.

We analyzed the statistical significant differences (using a one-way ANOVA) in individual search between two systems (i.e., S1 and S2) and among activities (i.e., Vsnip vs. Comt, Query, All Rel, and All View), at $p < 0.1$. The results showed better performance for S1 than that for S2 on average in Vsnip vs. Comt, All Rel, Query, and All View with statistical significance at $p = 0.035^*$, but not statistical significance at $p = 0.634$, $p = 0.188$, and $p = 0.966$, respectively.

A one-way ANOVA was also conducted to determine the statistical significant differences in asynchronous collaborative search between two systems (i.e., S1 and S2) and among activities (i.e., Vsnip with Comt, Query, All Rel, and All View), at $p < 0.1$. The results demonstrate better performance for S1 than that for S2 on average in Vsnip vs. Comt, All Rel, Query, and All View with statistical significance at $p = 0.095^*$ and $p = 0.004^*$, but not statistical significance at $p = 0.395$ and $p = 0.337$, respectively.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table 2. Average feedback from a post-task questionnaire using Five-Point-Likert scale (closer to 1 = disagree, closer to 5 = agree)

Question	System	Mean				Mean	S.D.
		T1	T2	T3	T4		
Q1. I feel that the search task was easy after finished it.	S1	**3.19**	**3.19**	2.13	**2.88**	**2.84**	1.12
	S2	2.63	2.31	**2.25**	2.69	2.47	1.42
Q2. I was relaxed while carrying out the search task.	S1	**3.94**	**3.81**	2.63	**3.52**	**3.52**	1.04
	S2	3.31	3.25	**3.31**	3.13	3.25	1.08
Q3. It was easy to formulate initial queries on these topics.	S1	**4.06**	**4.06**	**4.19**	**4.00**	**4.08**	0.55
	S2	3.63	3.88	3.94	3.88	3.83	0.72
Q4. I had enough time to do an effective search task.	S1	**4.06**	**4.19**	3.19	**3.88**	**3.83**	1.10
	S2	3.56	3.44	**3.63**	3.44	3.52	1.14
Q5. I have succeeded in my performance of the search task.	S1	**4.13**	**4.00**	3.38	**4.38**	**3.97**	0.91
	S2	3.13	3.38	**3.94**	3.75	3.55	0.87

5.2 User Feedbacks

Table 2 illustrates the average of user feedbacks in a post-task questionnaire. The participant feedbacks includes their satisfaction, opinion and their experience when using CoZpace with or without Snapboard. We analyzed the statistical significant differences (using a two-way ANOVA) between two systems (i.e., S1 and S2) and among search tasks (i.e., T1, T2, T3, and T4), at $p < 0.1$. The analysis shows better feedbacks for S1 than that for S2 on average in Q1, Q2, Q3, Q4, and Q5 with statistical significance at $p = 0.0568^*$, $p = 0.0856^*$, $p = 0.0829^*$, $p = 0.0934^*$, and $p = 0.0053^*$, respectively.

Moreover, as T1, T2, and T4 are ordinal interaction effect, the main effect can be considered. The main effect of system shows that the user's satisfaction on average in Q1, Q2, Q3, Q4, and Q5 are better for S1 than for S2 in T1, T2, and T4. There is no main effect of task among T1, T2, and T4 but there is a little bit effect from T3. We also tried to determine the statistical significant differences (using a two-way ANOVA) between two systems (i.e., S1 and S2) and among questionnaires (i.e., Q1, Q2, Q3, Q4, and Q5), at $p < 0.1$. The results show better feedbacks for S1 than that for S2 on average in T1, T2, T3 and T4 with statistical significance at $p = 0.0042^*$, $p = 0.0015^*$, $p = 0.0634^*$, and not statistical significance at $p = 0.1315$, respectively. As the tasks T1 and T2 basically require users to find information better presented in a visual form (i.e., pictures better convey the content toward collaborators than text), the results seem to be pertinent to our scenario of using Snapboard feature that it is appropriate for any task that needs to share pictures, such as sharing attractions for travel planning.

The difference in T3 is due to the fact that the literature search mostly related to finding text results might not take benefits from the Snapboard feature. As the difference in T4 is extremely not significant, the better feedbacks for S1 should not be claimed. However, it might be interpreted that the better feedbacks are not significant because T4 is a task that might not need to share pictures. The average feedbacks from Q1 in Table 2 can infer that all the four search tasks are quite difficult when using S2 but become easier when using S1 in T1, T2 and

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table 3. Average feedback about user satisfaction to our CoZpace system from an exit questionnaire using Five-Point-Likert scale (closer to 1 = disagree, closer to 5 = agree)

Question	Mean	S.D.
1. The search interface was easy to use.	4.03	0.82
2. Retrieval function was helpful in exploring web.	4.13	0.71
3. Comment function in S1 was helpful in exploring web.	3.34	1.24
4. Snapshot function in S2 was helpful in exploring web.	**4.41**	0.71

T4. This inference answers our RQ1 that Snapboard feature supports exploratory search tasks.

Table 3 shows quantitative data gathered from an exit questionnaire. Users were asked to rate their satisfaction to the interface of CoZpace as well as its three main features, i.e., retrieval, comment and snapshot functions. The satisfaction of Snapshot function on average is 4.41 with a low standard deviation showing a very high level of agreement. On the other hand, the comment feature got the lowest average score, 3.34. For the CoZpace interface, the results are very positive with a mean value of 4.03.

Upon the completion of study, we arranged a discussion session with participants. They provided open comments with respect to the feature of Snapboard. Some examples include: "Visual snippet helps me to easily share needed information with collaborators" and "Snapboard increases the awareness of how our tasks progress, and I can use it to identify relevant information in web pages or revisit them later". Most of the participants agree that those benefits can help them to collaboratively complete the search tasks easier.

6 Conclusion

In this paper, we conducted a user study to evaluate the supportiveness of Snapborad, i.e., visual snippets, in carrying out exploratory search tasks in two scenarios of individual and asynchronous collaborative search. We experimented on two variations of CoZpace with pairs of participants by collecting all activities logs and their feedbacks on the usage of the systems, in particular those related to Snapboard. The evaluation showed that Snapboard helped the participants in the collaborative search tasks, especially Travel planning and General shopping, in both scenarios. The reason is that most traveling and shopping web pages consist of not only texts but also several images which are easier to recognize and to describe in picture than in text format. Furthermore, according to the evaluation, using Snapboard can generate more relevant web pages. The number of relevant web pages in S1 is higher than that in S2 in both individual search and asynchronous collaborative search. Thus, it means that the Snapboard support and improve in both searching conditions.

For future work, we plan to conduct a user study to evaluate the effectiveness in synchronous collaborative search. In addition, we will compare the proposed our visual snippets with other kinds of summarization (e.g. thumbnails.)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

References

1. Amershi, S., Morris, M.R.: Cosearch: A system for co-located collaborative web search. In: Proceedings of CHI '08. pp. 1647–1656. Florence, Italy
2. Aula, A., Khan, R.M., Guan, Z., Fontes, P., Hong, P.: A comparison of visual and textual page previews in judging the helpfulness of web pages. In: Proceedings of WWW '10. pp. 51–60. Raleigh, USA
3. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Information Review 13(5), 407–424 (1989)
4. Borlund, P.: User-centered evaluation of information retrieval systems. Information Retrieval: searching in the 21st Century pp. 21–37 (2009)
5. Broder, A.: A taxonomy of web search. SIGIR Forum 36(2), 3–10 (Sep 2002)
6. Bruce, H., Jones, W., Dumais, S.: Keeping and re-finding information on the web: What do people do and what do they need? In: ASIST '04 (2004)
7. Golovchinsky, G., Pickens, J., Back, M.: A taxonomy of collaboration in online information seeking. In: Proceedings of the 1st International Workshop on Collaborative Information Retrieval. JCDL '08, Pittsburgh, USA
8. Halvey, M., Vallet, D., Hannah, D., Feng, Y., Jose, J.M.: An asynchronous collaborative search system for online video search. Information processing & management 46(6), 733–748 (2010)
9. Jiao, B., Yang, L., Xu, J., Wu, F.: Visual summarization of web pages. In: Proceedings of SIGIR '10. pp. 499–506. Geneva, Switzerland
10. Kruajirayu, H., Leelanupab, T.: Extracting visual snippets for query suggestion in collaborative web search. In: Proceedings of the 20th International Society on Artificial Life and Robotics. AROB 2015, Beppu, Japan (2015)
11. Kruajirayu, H., Tangsomboon, A., Leelanupab, T.: Cozpace: A proposal for collaborative web search for sharing search records and interactions. In: Student Project Conference (ICT-ISPC), 2014 Third ICT International. pp. 165–168. IEEE (2014)
12. Morris, M.R.: A survey of collaborative web search practices. In: Proceedings of CHI '08. pp. 1657–1660. Florence, Italy
13. Morris, M.R., Horvitz, E.: Searchtogether: An interface for collaborative web search. In: Proceedings of UIST '07. pp. 3–12. Newport, USA
14. Morris, M.R., Lombardo, J., Wigdor, D.: Wesearch: supporting collaborative search and sensemaking on a tabletop display. In: Proceedings of CSCW '10. pp. 401–410. Savannah, USA
15. Rodriguez Perez, J., Leelanupab, T., Jose, J.M.: Cofox: A synchronous collaborative browser. In: Proceedings of AIRS '12. pp. 262–274. Tianjin, China
16. Shah, C., Marchionini, G.: Awareness in collaborative information seeking. JASIST 61(10), 1970–1986 (2010)
17. Teevan, J., Cutrell, E., Fisher, D., Drucker, S.M., Ramos, G., André, P., Hu, C.: Visual snippets: Summarizing web pages for search and revisitation. In: Proceedings of CHI '09. pp. 2023–2032. Boston, USA
18. Turpin, A., Tsegay, Y., Hawking, D., Williams, H.E.: Fast generation of result snippets in web search. In: Proceedings of SIGIR '07. pp. 127–134. Amsterdam, Netherlands
19. White, R.W., Roth, R.A.: Exploratory Search: Beyond the Query-Response Paradigm. Synthesis Lectures on Information Concepts, Retrieval, and Services 1, 1–98 (2009)
20. Yue, Z., Han, S., He, D.: A comparison of action transitions in individual and collaborative exploratory web search. In: Proceedings of AIRS '12. pp. 52–63. Tianjin, China

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

เอกสารในการทำทดลองเรื่องการศึกษาระบบค้นหาข้อมูลแบบร่วมมือจากกลุ่มผู้ใช้งาน

ภาคผนวกนี้นำเสนอเอกสารในการทำทดลองเรื่องการศึกษาระบบค้นหาข้อมูลแบบร่วมมือจากกลุ่มผู้ใช้งาน ซึ่งได้ถูกอธิบายในบทที่ 4 ประกอบด้วยเอกสารดังต่อไปนี้

1. เอกสารชี้แจงผู้เข้าร่วมการทดลอง (Information Sheet)
2. แบบฟอร์มยินยอมเข้าร่วมการทดลอง (Consent Form)
3. เอกสารอธิบายโจทย์ที่ใช้ในการทดลอง (Task Descriptions)
4. แบบสอบถามก่อนทำการทดลอง (Entry Questionnaire)
5. แบบสอบถามหลังทำการทดลอง (Post-Search Questionnaire)
6. แบบสอบถามหลังสำเร็จการทดลอง (Exit Questionnaire)

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. เอกสารชี้แจงผู้เข้าร่วมการทดลอง (Information Sheet)

INFORMATION SHEET

Experiment: A study of users' interaction on search collaboration using CoZpace

Researcher: Hannarin Kruajirayu

You are invited to take part in a research study. Before you decide to do so, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask me if anything is not clear or if you would like more information.

The objective of this experiment is to evaluate and compare the relative effectiveness of two different collaborative search systems, i.e., collaborative search system with and without a *Snapboard* feature. The evaluation consists of requiring users to search, investigating their interactions with the systems, and asking them to answer the questionnaires. Please remember that they are the systems, not you, that are being evaluated. It is up to you to decide whether or not to take part. If you decide to take part, you will be given this information sheet to keep and asked to sign a consent form. You are free to withdraw at any time without giving a reason. You also have the right to withdraw retrospectively any consent given, and to require that any data gathered on you be destroyed.

The experiment will last around three hours. You will complete search tasks using two different search systems. You will be given a chance to learn how to use all the two systems before we begin. At this time you will also be asked to complete an entry questionnaire. You will perform four tasks in total. There is a time limit for each task which takes 20 minutes. After completing each task, you will be asked to fill in a questionnaire about your experience during the search and all of your interactions (e.g., chat messages, submitted queries, and results marked as relevant and taken snapshots of information) will also be logged. You are encouraged to comment on each interface as you use it. Please ask questions if you need to and please let me know when you are finished with the task. Finally, after completing all tasks, you will be asked some questions about the tasks, your search strategy and the systems. Remember, you can opt out at any time during the experiment.

All information collected about you during the course of this study will be kept strictly confidential. You will be identified by an ID number and all information about you will have your name and contact details removed so that you cannot be recognised from it. We confirm that all of the information gathered on you will be only used for this study, and then destroyed. For further information about this study please contact.

Hannarin Kruajirayu
Faculty of Information Technology, King Mongkuts Institute of Technology Ladkrabang (KMITL)
Chalongkrung Rd. Ladkrabang
Bangkok, Thailand. 10520
Email: hannarin_k@hotmail.com
Tel.: 08-5959-5938

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. แบบฟอร์มยินยอมเข้าร่วมการทดลอง (Consent Form)

CONSENT FORM

Project: **A study of users' interaction on search collaboration using CoZpace**

Researcher: **Hannarin Kruajirayu**

Please tick box

1. I confirm I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that my permission is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected. ☐
3. I agree to take part in the above study. ☐

Please leave your tel. number and email address for future contact

Tel. _____ Email _____

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Researcher	Date	Signature

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เอกสารอธิบายโจทย์ที่ใช้ในการทดลอง (Task Descriptions)

EXPERIMENT DESCRIPTION 1

Experiment: A study of users' interaction during the search process with asynchronous collaboration (AC)

Researcher: Hannarin Kruajirayu

User ID:

Topic: Travel planning

Simulated situation:

Simulative work task situation:

Imagine you and your friend are planning a trip to the Kansas City, USA. You want to search for information about how you will spend your vacation in USA. You will be there for a month.

Indicative request:

Your goal is to find, where you will stay, what you will do, and how will you get there, etc.

Description:

- *In all tasks, a participant (P1) starts the first session of the search tasks and finishes in 20 minutes.*
- *Afterwards, the participant (P2) continues the same search task of P1 and finishes in 20 minutes.*

This experiment is called "asynchronous collaborative search."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EXPERIMENT DESCRIPTION 2

Experiment: A study of users' interaction during the search process with asynchronous collaboration (AC)

Researcher: Hannarin Kruajirayu

User ID:

Topic: General shopping

Simulated situation:

Simulative work task situation:

Imagine you and your brother are given a budget to buy a car (10,000$). You want to search information about cars and the differences among each car.

Indicative request:

Your goal is to find the technical specifications of cars, brands of cars, or stores which sell cars, etc.

Description:

- *In all tasks, a participant (P1) starts the first session of the search tasks and finishes in 20 minutes.*
- *Afterwards, the participant (P2) continues the same search task of P1 and finishes in 20 minutes.*

This experiment is called "asynchronous collaborative search."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EXPERIMENT DESCRIPTION 3

Experiment: A study of users' interaction during the search process with asynchronous collaboration (AC)

Researcher: Hannarin Kruajirayu

User ID: ☐

Topic: Literature search

Simulated situation:

Simulative work task situation:

Imagine you and your classmate are assigned to write a summary article about the US civil war. You want to find general information and as many relevant documents.

Indicative request:

Your task is to find, causes of the civil war, economic causes, consequences of the civil war, civil war effects today, or weapons were used during civil war, etc.

Description:

- *In all tasks, a participant (P1) starts the first session of the search tasks and finishes in 20 minutes.*
- *Afterwards, the participant (P2) continues the same search task of P1 and finishes in 20 minutes.*

This experiment is called "asynchronous collaborative search."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

EXPERIMENT DESCRIPTION 4

Experiment: A study of users' interaction during the search process with asynchronous collaboration (AC)

Researcher: Hannarin Kruajirayu

User ID:

Topic: Technical information

Simulated situation:

Simulative work task situation:

Imagine you and your brother have decided that you want to reduce the use of air conditioning in your house. You've thought that if you could protect the roof being overly hot due to sun exposure, you could keep the house temperature low without the excessive use of air conditioning.

Indicative request:

Your task is to find, the best material to use if the purpose is cooling down the roof so that the house temperature remains low, or other solutions regarding roof coating, etc.

Description:

- *In all tasks, a participant (P1) starts the first session of the search tasks and finishes in 20 minutes.*
- *Afterwards, the participant (P2) continues the same search task of P1 and finishes in 20 minutes.*

This experiment is called "asynchronous collaborative search."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. แบบสอบถามก่อนทำการทดลอง (Entry Questionnaire)

ENTRY QUESTIONNAIRE

This questionnaire will provide us with background information that will help us analyse the answers you give in later stages of this experiment. You are not obliged to answer a question, if you feel it is too personal.

User ID:	

Please place a TICK ☑ in the square that best matches your opinion.

Part 1: PERSONAL DETAILS

1. Please provide your AGE:	

2. Please indicate your GENDER:			
Male	☐	Female	☐

3. Please provide your current OCCUPATION/STUDY:	

4. What is your FIELD of work or study?	

5. What is your educational level?			
Undergraduate/No Degree	1 ☐	Graduate Student/Primary Degree	2 ☐
Researcher/Advanced Degree	3 ☐	Faculty/Research Staff	4 ☐

5. How would you describe your proficiency with ENGLISH			
Native Speaker	1 ☐	Advanced	2 ☐
Intermediate	3 ☐	Beginner	4 ☐

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Part 2: SEARCH EXPERIENCE

1. How often do you search the Web?			
Never	1 ☐	Once or twice a year	2 ☐
Once or twice a month	3 ☐	Once or twice a week	4 ☐
Once or twice a day	5 ☐	More often	6 ☐

2. What is the Web Source Type, you often search?			
Web	1 ☐	Image	2 ☐
News	3 ☐	Video	4 ☐

3. Have you ever collaborated with other people to search the web?			
Yes	1 ☐	No	2 ☐

4. How do you interact with each other while searching?			
Instant messaging	1 ☐	Phone	2 ☐
Email	3 ☐	Videoconferencing	4 ☐
Other	5 ☐		

6. Have you ever used collaborative search system?			
Yes	1 ☐	No	2 ☐

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. แบบสอบถามหลังทำการทดลอง (Post-Search Questionnaire)

Post-Task 1 Questionnaire (Asynchronous collaboration)

To evaluate the system you have just used, we now ask you to answer some questions about it. Take into account that we are interested in knowing your opinion: answer questions freely, and consider there are no right or wrong answers. Please remember that we are evaluating the system you have just used and not you.

User ID:		System:		Task:	Travel planning

Please place a TICK ☑ in the square that best matches your opinion. Please answer all questions.

Part 1: TASK

In this section we ask about the search tasks you have just attempted.

1. The task we asked you to perform was:

easy	☐	☐	☐	☐	☐	difficult
stressful	☐	☐	☐	☐	☐	relaxing

2. It was easy to formulate initial queries on these topics.

Disagree				Agree
☐	☐	☐	☐	☐
1	2	3	4	5

3. I had enough time to do an effective search.

Disagree				Agree
☐	☐	☐	☐	☐
1	2	3	4	5

4. I believe I have succeeded in my performance of the task.

Disagree				Agree
☐	☐	☐	☐	☐
1	2	3	4	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. แบบสอบถามหลังสำเร็จการทดลอง (Exit Questionnaire)

Exit Questionnaire/Interview

The aim of this experiment was to investigate the effectiveness of two different collaborative search systems. Please consider the entire search experience that you just had when you respond to the following questions.

User ID:	

Please place a TICK ☑ in the square that best matches your opinion. Please answer the questions as fully as you feel able to.

Question	Disagree				Agree
	1	2	3	4	5
1. The search interface was not difficult to use.					
2. The system return relevant document.					
3. I found, the system response time was fast.					
4. Relevant feature was helpful in exploring web.					
5. Snapshot feature was helpful in exploring web.					
6. Comment feature was helpful in exploring web.					

Which systems did you ...	CoZpace with Snapshot	CoZpace Without Snapshot	No difference
7. ... find BEST overall?			
8. ... find easier to LEARN TO USE?			
9. ... find easier to USE?			
10. ... PERFER?			
11. ... find changed your perception of the task?			
12. ... find more EFFECTIVE for the tasks you performed?			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

13. What did you LIKE about each of the systems?
System 1 (CoZpace with Snapshot) :
System 2 (CoZpace without Snapshot) :

14. What did you DISLIKE about each of the systems?
System 1 (CoZpace with Snapshot) :
System 2 (CoZpace without Snapshot) :

15. Additional Comments (Optional)?

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ผู้เขียน	นายหาญนรินทร์ เครือจิรายุส์
วันเดือนปีเกิด	10 สิงหาคม 2534
สถานที่เกิด	จังหวัด ราชบุรี
ปริญญา	2557 วิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง 2558 วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงเทคโนโลยีระบบสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
งานวิจัยที่ตีพิมพ์	*การพัฒนาระบบค้นคืนสารสนเทศสำหรับการร่วมมือค้นหาข้อมูลแบบกลุ่ม* **Hannarin Kruajirayu**, Ake Tangsomboon, and Teerapong Leelanupab; in Proceedings of the NCIT 2014, Thailand. *CoZpace: A Proposal for Collaborative Web Search for Sharing Search Records and Interactions.* **Hannarin Kruajirayu**, Ake Tangsomboon, and Teerapong Leelanupab; in Proceedings of the ICT-ISPC 2014. Thailand. *Extracting Visual Snippets for Query Suggestions in Collaborative Web Search.* **Hannarin Kruajirayu**, and Teerapong Leelanupab; in Proceedings of the AROB 2015, Beppu, Japan.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Snapboard: A Shared Space of Visual Snippets. A Study in Individual and Asynchronous Collaborative Web Search.

Teerapong Leelanupab, **Hannarin Kruajirayu**, and Nont Kanungsukkasem; in Proceedings of the AIRS 2015, Brisbane, Queensland, Australia.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้