ระบบแนะนำโดยใช้แท็กคลาวด์เสมือน

นายคุณารักษ์ โอสถาภิรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมซอฟต์แวร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository(CUIR)
are the thesis authors' files submitted through the Graduate School.

RECOMMENDER SYSTEMS USING VIRTUAL TAG CLOUD

Mr. Kunaruk Osatapirat

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Software Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ระบบแนะนำโดยใช้แท็กคลาวด์เสมือน
โดย	นายคุณารักษ์ โอสถาภิรัตน์
สาขาวิชา	วิศวกรรมซอฟต์แวร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

…… ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์)

…… กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต)

คุณารักษ์ โอสถาภิรัตน์: ระบบแนะนำโดยใช้แท็กคลาวด์เสมือน.
(RECOMMENDER SYSTEMS USING VIRTUAL TAG CLOUD)
อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.ญาใจ ลิ่มปิยะกรณ์, 72 หน้า.

ผู้ใช้งานในสังคมออนไลน์ใช้โซเชียลแท็กกิ้งสำหรับกำกับแหล่งข้อมูลด้วยคำอิสระในรูปแบบของเมทาดาตา ซึ่งอธิบายลักษณะหรือความสนใจของผู้ใช้งาน อย่างไรก็ตาม ระบบแนะนำส่วนใหญ่ในปัจจุบันได้ใช้พฤติกรรมการแท็กของผู้ใช้งานเพื่อทำการแนะนำเนื้อหาโดยมุ่งเน้นที่แท็กที่ผู้ใช้งานระบุซึ่งเป็นความสนใจทางตรงของผู้ใช้งานเท่านั้น งานวิจัยนี้ได้นำเสนอแนวทางในการดักจับความสนใจทั้งทางตรงและทางอ้อมโดยใช้แท็กคลาวด์เสมือนหรือวีทีซีสำหรับระบบแนะนำอิงการติดแท็ก การสร้างแท็กคลาวด์เสมือนและแบบจำลองการแนะนำแหล่งข้อมูลโดยการวิเคราะห์จากแท็กคลาวด์เสมือนได้ถูกอธิบายไว้ในงานนี้ ประสิทธิผลของวิธีการที่นำเสนอถูกประเมินด้วยค่าตัววัดเอฟ

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....
สาขาวิชา วิศวกรรมซอฟต์แวร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....
ปีการศึกษา 2554 .....

5270736321 : MAJOR SOFTWARE ENGINEERING

KEYWORDS : RECOMMENDER SYSTEM / SOCIAL TAGGING / VIRTUAL TAG CLOUD

KUNARUK OSATAPIRAT : RECOMMENDER SYSTEMS USING VIRTUAL TAG CLOUD. ADVISOR : ASSOC. PROF. YACHAI LIMPIYAKORN, Ph.D. 72 pp.

Users in social network use social tagging to annotate the free text to the resource in terms of metadata, which describes users' characteristics or interests. However, the majority of current recommender systems use the users' tagging behaviors to recommend the resource by focusing on what users tag or direct interests only. This research presents an approach to capturing user direct and indirect interests using Virtual Tag Cloud (VTC) for the tag-based recommender systems. The manipulation of VTC and the resource recommendation model are described in this work. The assessment of the effectiveness of the proposed method is measured by F-Measure.

Department : Computer Engineering Student's Signature .....

Field of Study : Software Engineering Advisor's Signature .....

Academic Year : 2011

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความอนุเคราะห์อย่างดียิ่งของรองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้สละเวลาให้ความรู้ ให้คำปรึกษา ตรวจสอบ ให้คำแนะนำแนวทางการวิจัย และสนับสนุน จนทำให้การวิจัยในครั้งนี้สำเร็จออกมาด้วยดี ข้าพเจ้าจึงขอกราบระลึกถึงพระคุณของอาจารย์ไว้ ณ ที่นี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ และอาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต กรรมการสอบวิทยานิพนธ์ ที่กรุณาเสียสละเวลา ให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้

ท้ายที่สุด ผู้เสนอวิทยานิพนธ์ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัว สำหรับกำลังใจที่มีค่ายิ่ง รวมถึงขอขอบพระคุณผู้บังคับบัญชาในสายงาน เพื่อนร่วมงาน และมิตรสหาย ที่คอยติดตามให้กำลังใจ ให้การสนับสนุนและความช่วยเหลือในด้านต่างๆ และท่านอื่นๆ ที่มิได้กล่าวชื่อไว้ ณ ที่นี้ที่มีส่วนช่วยให้วิทยานิพนธ์ของข้าพเจ้าสำเร็จไปได้ด้วยดี

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ฎ

บทที่ 1 บทนำ 15

1.1 ความเป็นมาและความสำคัญของปัญหา 15

1.2 วัตถุประสงค์ของการวิจัย 16

1.3 ขอบเขตของการวิจัย 16

1.4 ประโยชน์ที่คาดว่าจะได้รับ 16

1.5 วิธีดำเนินงานวิจัย 16

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์ 17

1.7 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์ 17

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 18

2.1 ทฤษฎีที่เกี่ยวข้อง 18

2.2 งานวิจัยที่เกี่ยวข้อง 21

บทที่ 3 วิธีดำเนินการวิจัย 26

3.1 แนวคิดในการพัฒนา 26

3.2 การจัดการแท็กคลาวด์เสมือนและการปรับน้ำหนัก 27

3.3 การแนะนำแหล่งข้อมูล 29

บทที่ 4 การออกแบบและพัฒนาระบบ 31

4.1 สถาปัตยกรรมระบบ 31

4.2 สภาพแวดล้อมและเครื่องมือที่ใช้ในการพัฒนา 36

4.3 การพัฒนาระบบ 36

บทที่ 5 การประเมินและการวัดผล 46

5.1 วิธีการประเมินความแม่นยำของการแนะนำแหล่งข้อมูล 46

หน้า

5.2 การเตรียมข้อมูลเพื่อทำการทดสอบ 47
5.3 การปรับค่าพารามิเตอร์.....47
5.4 ผลการทดลอง 48
บทที่ 6 สรุปผลการวิจัย และข้อเสนอแนะ 64
6.1 สรุปผลการวิจัย 64
6.2 ข้อจำกัด 65
6.3 แนวทางการวิจัยต่อ 65
รายการอ้างอิง.....67
ภาคผนวก.....68
ประวัติผู้เขียนวิทยานิพนธ์.....72

สารบัญตาราง

หน้า

ตารางที่ 1 โครงสร้างข้อมูลที่จัดเก็บข้อมูลผู้ใช้บริการ (Author) 39
ตารางที่ 2 โครงสร้างข้อมูลที่จัดเก็บแหล่งข้อมูล (Item) 40
ตารางที่ 3 โครงสร้างข้อมูลแท็กที่อยู่ใน Author และ Item 40
ตารางที่ 4 ตัวอย่างผลการคำนวณค่า Precision, Recall และ F-Measure 69
ตารางที่ 5 ตัวอย่างผลการคำนวณค่า Precision เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด . 70
ตารางที่ 6 ตัวอย่างผลการคำนวณค่า Recall เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด 70
ตารางที่ 7 ตัวอย่างผลการคำนวณค่า F-Measure เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด
..... 71

สารบัญภาพ

หน้า

ภาพที่ 1 การใช้แท็กในสังคมออนไลน์ 18
ภาพที่ 2 ตัวอย่างแท็กคลาวด์จากเว็บไซต์ Delicious [2] 19
ภาพที่ 3 ตัวอย่างแท็กคลาวด์จากเว็บไซต์ Flickr [3] 19
ภาพที่ 4 การกรองร่วม [4] 20
ภาพที่ 5 แผนภูมิแสดงสัดส่วนพฤติกรรมการใช้แท็กของผู้ใช้งานในสังคมออนไลน์ [5] 21
ภาพที่ 6 ภาพรวมของระบบแนะนำโดยใช้ข้อมูลแท็กและเวลา [6] 22
ภาพที่ 7 ค่าความแม่นยำจากระบบที่วิเคราะห์และแนะนำจากค่าน้ำหนักในรูปแบบต่างๆ [6] 23
ภาพที่ 8 แผนภูมิแสดงจำนวนบทความที่แนะนำตามขนาดของข้อมูล [7] 24
ภาพที่ 9 แผนภูมิแสดงค่าความแม่นยำของสมการความคล้ายในสภาพแวดล้อมของระบบที่ทำการทดลอง [8] 25
ภาพที่ 10 ภาพรวมการทำงานของระบบ 26
ภาพที่ 11 การสร้างแท็กคลาวด์เสมือนของแหล่งข้อมูลเมื่อมีการเพิ่มแหล่งข้อมูลเข้าสู่ระบบ 28
ภาพที่ 12 เพิ่มแท็กในแท็กคลาวด์เสมือนเมื่อมีการเพิ่มแท็กให้กับแหล่งข้อมูล 28
ภาพที่ 13 การเพิ่มแท็กในแท็กคลาวด์เสมือนเมื่อมีการแนะนำต่อ 28
ภาพที่ 14 สถาปัตยกรรมระบบ 31
ภาพที่ 15 คลาสไดอะแกรมของส่วนที่ดึงข้อมูลจากเว็บไซต์ Delicious.com 32
ภาพที่ 16 แผนภาพลำดับเหตุการณ์ของส่วนดึงข้อมูลจากเว็บไซต์ Delicious.com 32
ภาพที่ 17 คลาสไดอะแกรมของส่วนที่จำลองการใช้งานของผู้ใช้งานเว็บไซต์ Delicious.com 33
ภาพที่ 18 แผนภาพลำดับเหตุการณ์ของส่วนที่จำลองการใช้งานของผู้ใช้งานเว็บไซต์ Delicious.com 34
ภาพที่ 19 คลาสไดอะแกรมของส่วนของการแนะนำแหล่งข้อมูลและเชื่อมต่อผู้ใช้งาน 35
ภาพที่ 20 แผนภาพลำดับเหตุการณ์ของส่วนแนะนำแหล่งข้อมูลและเชื่อมต่อผู้ใช้งาน 35
ภาพที่ 21 โครงสร้างของข้อมูล XML ที่โหนด rss 37
ภาพที่ 22 โครงสร้างของข้อมูล XML ที่โหนด channel 37
ภาพที่ 23 โครงสร้างของข้อมูล XML ที่โหนด Item 38
ภาพที่ 24 ตัวอย่างข้อมูลที่ได้จากเว็บไซต์ Delicious.com 38
ภาพที่ 25 ตัวอย่างการเรียกดูข้อมูลแหล่งข้อมูลจากตาราง Item 41

ภาพที่ 26 หน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการ 42
ภาพที่ 27 หน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการโดยเลือกวิธีคำนวณโดยใช้แท็คลาวด์เสมือน . 42
ภาพที่ 28 หน้าจอระบุชื่อเข้าแสดงแหล่งข้อมูลที่ระบบแนะนำ 43
ภาพที่ 29 หน้าจอแสดงแผนภูมิการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์หรือแท็กคลาวด์เสมือนของผู้ใช้บริการ 44
ภาพที่ 30 หน้าจอแสดงข้อมูลแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนของแหล่งข้อมูลที่ระบบแนะนำ 44
ภาพที่ 31 หน้าจอแสดงแผนภูมิการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์หรือแท็กคลาวด์เสมือนของแหล่งข้อมูล 45
ภาพที่ 32 หน้าจอแสดงรายการแหล่งข้อมูลที่ระบบแนะนำโดยใช้การคำนวณจากแท็กคลาวด์แบบปกติ 49
ภาพที่ 33 หน้าจอแสดงรายการแหล่งข้อมูลที่ระบบแนะนำโดยใช้การคำนวณจากแท็กคลาวด์เสมือน 49
ภาพที่ 34 ตำแหน่งของแหล่งข้อมูลที่แนะนำที่แตกต่างไปเมื่อใช้การคำนวณจากแท็กคลาวด์เสมือน 50
ภาพที่ 35 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล . 50
ภาพที่ 36 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล . 51
ภาพที่ 37 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล . 51
ภาพที่ 38 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของผู้ใช้บริการ .. 51
ภาพที่ 39 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานแบบปกติ 52
ภาพที่ 40 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานแบบปกติ 52
ภาพที่ 41 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานบ่อย 53
ภาพที่ 42 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนจากผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานบ่อย 53
ภาพที่ 43 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานมากกว่าปกติ 54

ภาพที่ 44 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานมากกว่าปกติ 54
ภาพที่ 45 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการเดียวกันที่ค่าขีดแบ่งเท่ากับ 0.03 55
ภาพที่ 46 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการเดียวกันที่ค่าขีดแบ่งเท่ากับ 0.03 55
ภาพที่ 47 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 56
ภาพที่ 48 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 57
ภาพที่ 49 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 57
ภาพที่ 50 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 58
ภาพที่ 51 แผนภูมิแสดงการเปรียบเทียบค่า Precision ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน 58
ภาพที่ 52 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 59
ภาพที่ 53 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 59
ภาพที่ 54 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 60
ภาพที่ 55 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 60
ภาพที่ 56 แผนภูมิแสดงการเปรียบเทียบค่า Recall ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน 61
ภาพที่ 57 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 61

หน้า

ภาพที่ 58 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ 62
ภาพที่ 59 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 62
ภาพที่ 60 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน 63
ภาพที่ 61 แผนภูมิแสดงการเปรียบเทียบค่า F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน 63

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบัน โซเชียลแท็กกิ้ง (Social tagging) หรือการระบุเมทาดาตา (metadata) ให้กับแหล่งข้อมูลที่ได้มีการแบ่งปันผ่านทางสังคมออนไลน์นั้น ได้มีการใช้งานกันอย่างแพร่หลายในเว็บสังคมออนไลน์สมัยใหม่ซึ่งอนุญาตให้ผู้ใช้บริการสามารถระบุแท็กที่สนใจไว้กับข้อมูลส่วนตัวหรือระบุแท็กให้กับแหล่งข้อมูลต่างๆ ในเว็บไซต์ได้ ทั้งนี้ เมื่อสังคมออนไลน์มีการเติบโตและมีขนาดของฐานข้อมูลที่ใหญ่มากขึ้น ข้อมูลแท็กที่มีอยู่ในระบบก็เปรียบเสมือนแหล่งข้อมูลหรือองค์ความรู้ที่สามารถนำมาใช้ในการสืบค้น รวมไปถึงการจัดกลุ่มข้อมูลและวิเคราะห์ความต้องการของผู้ใช้บริการที่สามารถนำมาพัฒนาต่อยอดเพื่อเพิ่มศักยภาพในการให้บริการ

จุดเด่นของการทำโซเชียลแท็กกิ้ง คือ การให้อิสระกับผู้ใช้บริการในการระบุคุณลักษณะของข้อมูลที่มีการแบ่งปันร่วมกันบนสังคมออนไลน์ ซึ่งทำให้มีความยืดหยุ่นสูงจนส่งผลให้เกิดความหลากหลายในเชิงบริบทของข้อมูล อย่างไรก็ตาม ด้วยความเป็นอิสระของข้อมูลทำให้ข้อมูลมีจำนวนมากและมีความหลากหลายจนเกินความจำเป็น มีข้อมูลที่ไม่สามารถนำมาใช้งานได้เป็นจำนวนมาก ทำให้ไม่สามารถนำข้อมูลแท็กทั้งหมดมาใช้ในการวิเคราะห์และคำนวณเพื่อแนะนำแหล่งข้อมูลที่น่าสนใจได้ตรงกับความต้องการของผู้ใช้บริการ

นอกจากปัญหาทางด้านจำนวนและความหลากหลายของแท็กแล้ว พบว่าการใช้ข้อมูลแท็กผู้ใช้บริการระบุอาจจะไม่ได้ระบุถึงความสนใจรองหรือความสนใจแฝงของผู้ใช้บริการ ดังตัวอย่างเช่น หากผู้ใช้บริการมีความสนใจในอาหารอเมริกันและญี่ปุ่น แต่มีการระบุแท็กที่ติดอยู่กับข้อมูลส่วนตัวว่า ”Hamburger” เพียงอย่างเดียว ซึ่งในความเป็นจริง ผู้ใช้บริการมีความสนใจเรื่องเกี่ยวกับ “Sushi” อีกด้วย ทั้งนี้ หากระบบทำการวิเคราะห์ข้อมูลแท็กจากข้อมูลส่วนตัวของผู้ใช้บริการเพียงอย่างเดียวก็จะแนะนำเฉพาะข้อมูลเกี่ยวกับ “Hamburger” เท่านั้น ทำให้ไม่สามารถตอบสนองความต้องการโดยรวมของผู้ใช้บริการได้

ผู้วิจัยได้เล็งเห็นถึงความสำคัญของความสนใจแฝงที่ผู้ใช้บริการไม่ได้ทำการระบุไว้กับข้อมูลส่วนตัว จึงได้นำเสนอวิธีการเก็บข้อมูลและวิเคราะห์ข้อมูลแท็กที่อยู่ในเครือข่ายสังคมออนไลน์ โดยใช้วิธีการสร้างแท็กคลาวด์เสมือน (Virtual Tag Cloud) ให้กับผู้ใช้บริการและแหล่งข้อมูลที่มีอยู่ในระบบ ซึ่งสามารถปรับเปลี่ยนแท็กข้อมูลภายในคลาวด์ได้ตามพฤติกรรมการใช้งานของผู้ใช้บริการในระบบ เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูลและสามารถแนะนำแหล่งข้อมูลให้ตรงกับความสนใจหรือความต้องการโดยรวมของผู้ใช้บริการ

1.2 วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอวิธีการในการพัฒนาระบบแนะนำการสืบค้นโดยการสร้างแท็กคลาวด์เสมือนสำหรับผู้ใช้บริการและแหล่งข้อมูล เพื่อใช้วิเคราะห์ถึงความต้องการหรือความสนใจแฝงของผู้ใช้บริการและนำมาเป็นส่วนร่วมในการคำนวณสำหรับแนะนำแหล่งข้อมูลให้กับผู้ใช้บริการ

1.3 ขอบเขตของการวิจัย

1. ระบบพัฒนาขึ้นโดยใช้ฐานข้อมูลจากเว็บไซต์สังคมออนไลน์ Delicious สำหรับการสร้างแบบจำลองการสร้างแท็กคลาวด์เสมือนเพื่อใช้ในการทดสอบสมมติฐาน
2. ระบบที่พัฒนาขึ้นจะประกอบด้วยกระบวนการการวิเคราะห์และคำนวณเพื่อแนะนำผลจากแท็กคลาวด์จริงและเสมือนเท่านั้น แต่จะไม่รวมไปถึงการตีความหมายของแท็กข้อมูลหรือการแก้ไขแท็กที่มีข้อมูลที่ผิดพลาด
3. ระบบที่พัฒนาขึ้นจะจัดทำส่วนเชื่อมต่อผู้ใช้งานในรูปแบบของเว็บแอปพลิเคชันเพื่อแสดงผลการแนะนำให้กับผู้ใช้ระบบ โดยผู้ใช้ระบบจะตรงกับผู้ใช้งานของเว็บไซต์สังคมออนไลน์ Delicious ตามช่วงระยะเวลาที่ได้นำข้อมูลมาทำการวิจัย

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แนวทางวิธีการใหม่สำหรับการพัฒนาระบบแนะนำจากข้อมูลแท็กในสังคมออนไลน์ ซึ่งเพิ่มความแม่นยำของการแนะนำแหล่งข้อมูลให้สอดคล้องกับความต้องการหรือความสนใจจริงของผู้ใช้บริการ อีกทั้งเป็นการเพิ่มศักยภาพทางด้านการแข่งขันและเปิดช่องทางใหม่ๆ ให้กับการตลาดผ่านทางเครือข่ายสังคมออนไลน์

1.5 วิธีดำเนินงานวิจัย

1. รวบรวมข้อมูลจากเว็บไซต์สังคมออนไลน์เพื่อนำมาใช้เป็นข้อมูลในการสร้างแบบจำลองและทดสอบ
2. ศึกษากระบวนการสร้างแท็กคลาวด์เสมือนและการปรับค่าน้ำหนักตามพฤติกรรมของผู้ใช้บริการ
3. ออกแบบการทดลองและการวัดผล
4. ออกแบบระบบเพื่อสร้างแบบจำลองระบบแนะนำโดยใช้แท็กคลาวด์เสมือน
5. ศึกษาและออกแบบส่วนเชื่อมต่อผู้ใช้งานผู้แสดงผลการแนะนำในรูปแบบของเว็บแอปพลิเคชัน
6. พัฒนาระบบแนะนำโดยใช้แท็กคลาวด์เสมือน

7. พัฒนาระบบส่วนเชื่อมต่อผู้ใช้งาน
8. ทำการทดลอง เก็บข้อมูลและทำการวัดผล
9. สรุปผลการทดลอง และตีพิมพ์ผลงานทางวิชาการ
10. จัดทำวิทยานิพนธ์

1.6 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

วิทยานิพนธ์นี้แบ่งเนื้อหาออกเป็น 6 บท ดังต่อไปนี้ บทที่ 1 เป็นบทนำกล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ของการวิจัย ขอบเขตของการวิจัย ประโยชน์ที่คาดว่าจะได้รับและผลงานตีพิมพ์ บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง บทที่ 3 กล่าวถึงวิธีดำเนินการวิจัย บทที่ 4 กล่าวถึง การออกแบบและพัฒนาระบบตามแนวทางการวิจัยที่นำเสนอ บทที่ 5 กล่าวถึงวิธีการประเมินและวัดผลการทดลองและบทที่ 6 สรุปผลการวิจัย ข้อเสนอแนะและแนวทางสำหรับการวิจัยต่อในอนาคต

1.7 ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้นำเสนอบทความวิชาการในหัวข้อเรื่อง “Capturing Personal Direct and Indirect Interests in Social Tagging Systems Using Virtual Tag Space”., K.Osatapirat and Y.Limpiyakorn. International Conference of Future Technology (ICFIT 2011), Singapore, Sep 16-18, 2011, ISSN: 2010-460X, Volume 13, pp. 249-253

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นบทความวิชาการในหัวข้อเรื่อง “ระบบแนะนำโดยใช้แท็กคลาวด์เสมือน” โดย คุณารักษ์ โอสถาภิรัตน์ และญาใจ ลิ่มปิยะกรณ์ ในวารสารรามคำแหง ฉบับวิศวกรรมศาสตร์ (Ramkhamhaeng Journal of Engineering) ปีที่ ๖ ฉบับที่ ๑

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

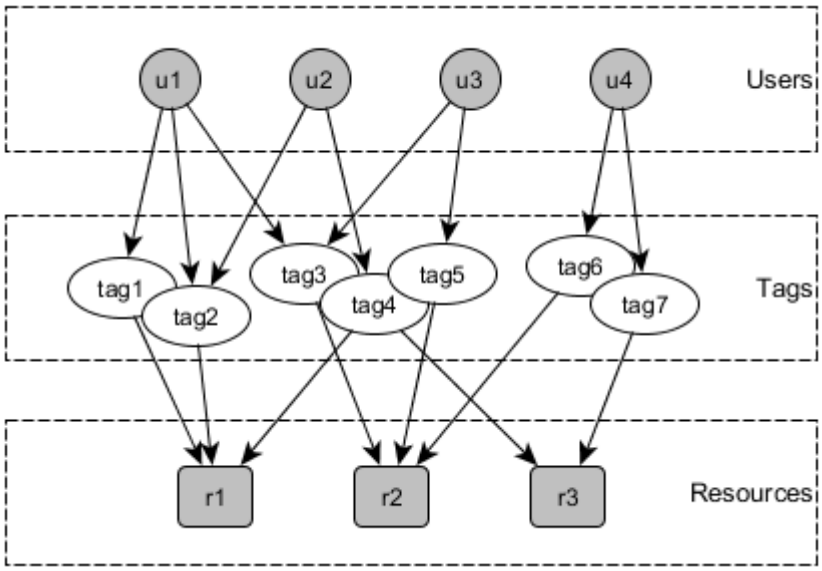
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 แท็กกิ้ง (Tagging)

แท็กกิ้ง คือ วิธีการที่ผู้ใช้บริการบนเครือข่ายอินเทอร์เน็ตใช้ในการจัดการการจัดหมวดหมู่หรือประเภทของข้อมูล สิ่งของ หรือวัตถุที่สนใจบนเครือข่ายอินเทอร์เน็ตหรือสังคมออนไลน์ โดยลักษณะข้อมูลที่ใช้โดยทั่วไปจะเป็นข้อความหรือเนื้อหาอิสระ (Free text) ซึ่งอธิบายถึงเมทาเดตาหรือรายละเอียดของวัตถุนั้นๆ

ข้อดีของการแท็กกิ้ง คือ การให้อิสระกับผู้ใช้บริการในการสร้างข้อมูล หรือผู้ใช้คือผู้ผลิต (User as a producer) ทำให้มีความยืดหยุ่นและรองรับการเพิ่มแท็กใหม่ๆ อยู่เสมอ และแท็กยังสามารถถูกนาเอามาใช้เป็นคาสืบค้นได้ ข้อเสียของการใช้แท็ก คือ ความผิดพลาดของข้อมูลมีโอกาสเกิดขึ้นค่อนข้างสูง เนื่องจากไม่สามารถตรวจสอบความถูกต้องของข้อมูล ณ ขณะเวลาที่ผู้ใช้บริการสร้างแท็กได้

การใช้งานแท็กในสังคมออนไลน์ในปัจจุบันนั้น โดยทั่วไปแล้วผู้ใช้บริการจะทำการระบุแท็กให้กับแหล่งข้อมูลโดยขึ้นอยู่กับบริบทที่ผู้ใช้บริการคิดว่าเหมาะสมกับแหล่งข้อมูลนั้นๆ ดังภาพที่ 1 เห็นได้ว่าข้อมูลแท็กหนึ่งๆ สามารถมีความสัมพันธ์กับแหล่งข้อมูลได้มากกว่าหนึ่งแหล่ง และในหนึ่งแหล่งข้อมูลนั้นก็สามารถมีการระบุแท็กได้หลายแท็กเช่นกัน

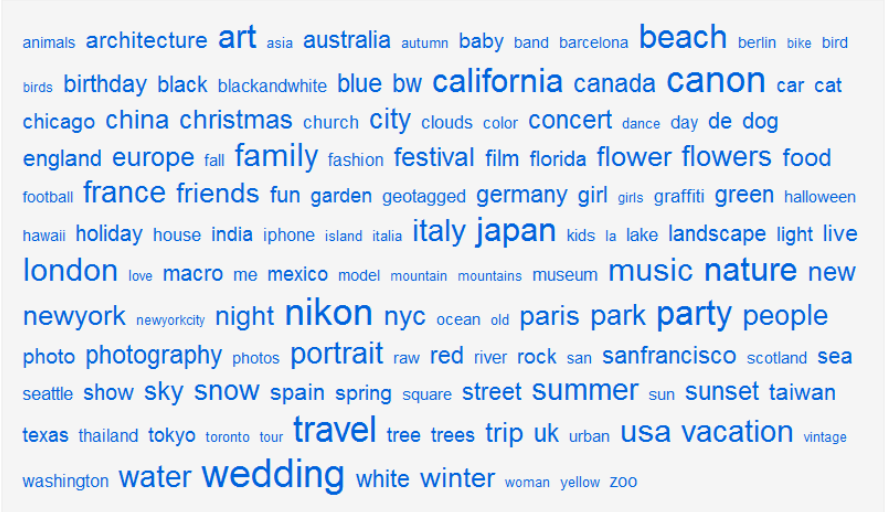


ภาพที่ 1 การใช้แท็กในสังคมออนไลน์

2.1.2 **แท็กคลาวด์** (Tag Cloud) [1]

แท็กคลาวด์ คือ การนำข้อมูลแท็กที่ติดอยู่กับผู้ใช้บริการหรือแหล่งข้อมูลมานำเสนอในรูปแบบที่แสดงลำดับความสำคัญหรือน้ำหนักของแท็กในระบบ โดยแท็กข้อมูลในแท็กคลาวด์นั้นจะมีการบ่งบอกถึงน้ำหนักความสำคัญของแท็ก ด้วยการใช้ขนาดตัวอักษรหรือสีที่แตกต่างกัน กล่าวคือ ขนาดตัวอักษรที่ใหญ่กว่าหรือสีที่เข้มกว่าบ่งบอกถึงน้ำหนักความสำคัญที่มีมากกว่าแท็กที่มีตัวอักษรเล็กหรือสีจางกว่า ประโยชน์ของการใช้แท็กคลาวด์ คือ สามารถแสดงให้เห็นถึงแท็กที่มีการใช้งานกันเป็นจำนวนมากในระบบ ตัวอย่างของแท็กคลาวด์จะมีลักษณะเป็นกลุ่มของข้อความ ดังภาพที่ 2 และ 3

ภาพที่ 2 ตัวอย่างแท็กคลาวด์จากเว็บไซต์ Delicious [2]



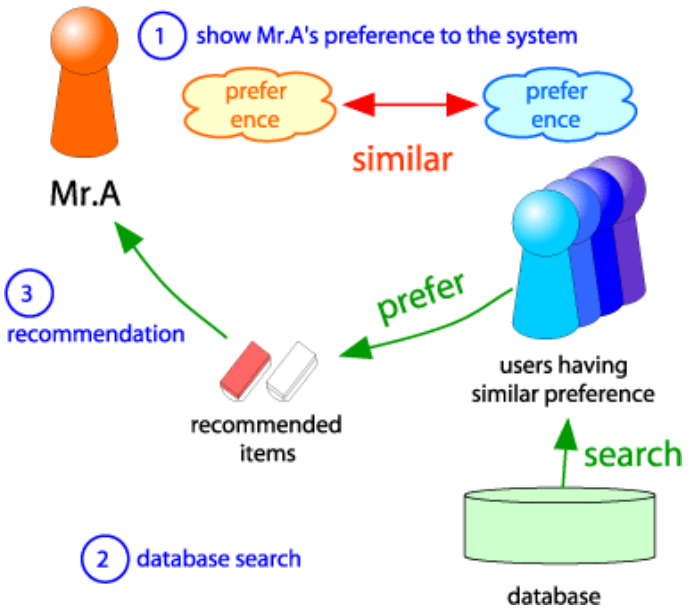
ภาพที่ 3 ตัวอย่างแท็กคลาวด์จากเว็บไซต์ Flickr [3]

ระบบที่จะพัฒนาขึ้นจะใช้วิธีการสร้างแท็กคลาวด์เสมือนขึ้นมาในระบบ โดยจะนำมาใช้ในการวิเคราะห์ความต้องการหรือความสนใจแฝงของผู้ใช้บริการ และนำมาคำนวณเพื่อเพิ่มความแม่นยำของการแนะนำแหล่งข้อมูลที่ผู้ใช้บริการมีความสนใจ

2.1.3 **การกรองร่วม** (Collaborative Filtering)

การกรองร่วมเป็นเทคนิคที่ใช้กันอย่างแพร่หลายเพื่อคัดกรองเอาเฉพาะข้อมูลที่น่าสนใจในระบบที่มีคุณสมบัติในการล่วงรู้บริบท (Context-Aware System) การกรองร่วมจะใช้หลักการการวิเคราะห์ข้อมูลจากหลายๆ ปัจจัย เช่น การเก็บรวบรวมข้อมูลคุณลักษณะของผู้ใช้บริการในระบบสำหรับการจัดกลุ่มผู้ใช้บริการที่มีลักษณะที่คล้ายคลึงกัน เพื่อนำไปวิเคราะห์หารายการสินค้าที่ตรงกับความต้องการของผู้ใช้บริการกลุ่มนี้ได้ ดังนั้น เมื่อมีผู้ใช้บริการรายใหม่ที่เพิ่งสมัครเข้ามาใช้บริการระบบก็จะทำให้สามารถทำนายได้ว่าผู้ใช้บริการต้องการสินค้าหรือบริการประเภทใด สำหรับแนวทางอื่นๆ อาจจะมีการใช้วิธีการทำเหมืองข้อมูล (Data Mining) ในการจดจำรูปแบบของการเลือกซื้อสินค้าโดยการทำ Association Mining หรือ Sequential Rules เพื่อให้การแนะนำผลการสืบค้นมีความแม่นยำและตรงกับความพึงพอใจของผู้ใช้บริการมากยิ่งขึ้น

ตัวอย่างเช่นภาพที่ 4 ระบบแนะนำสินค้าใช้การกรองร่วมในการแนะนำสินค้าให้กับ Mr.A โดยพิจารณาจากคุณลักษณะของ Mr.A และทำการค้นหาผู้ใช้งานอื่นๆ ในระบบที่มีคุณลักษณะที่คล้ายคลึงกับ Mr.A จากนั้นจึงทำการค้นหาสินค้าที่ผู้ใช้งานที่มีความคล้ายกับเคยเข้าไปเลือกซื้อหรือเลือกชมเพื่อนำมาแนะนำให้กับ Mr.A จากวิธีการข้างต้นจะเห็นว่ามีการใช้การกรองร่วมมาใช้เพื่อกรองเอาเฉพาะสินค้าที่คาดว่า Mr.A น่าจะสนใจเท่านั้นโดยใช้สมมติฐานที่ว่าผู้ใช้งานอื่นๆ ที่มีความคล้ายกับ Mr.A เคยเข้าไปเลือกซื้อหรือเลือกชมสินค้าเหล่านั้นแทนที่จะทำการวิเคราะห์ข้อมูลจากข้อมูลสินค้าทั้งระบบซึ่งมีปริมาณค่อนข้างมาก



ภาพที่ 4 การกรองร่วม [4]

ข้อดีของการทำการกรองร่วมคือ มีความแม่นยำค่อนข้างสูง เนื่องจากคำนวณจาก

คุณลักษณะในหลายๆ มิติของทั้งผู้ใช้บริการและวัตถุหรือสิ่งที่สนใจ

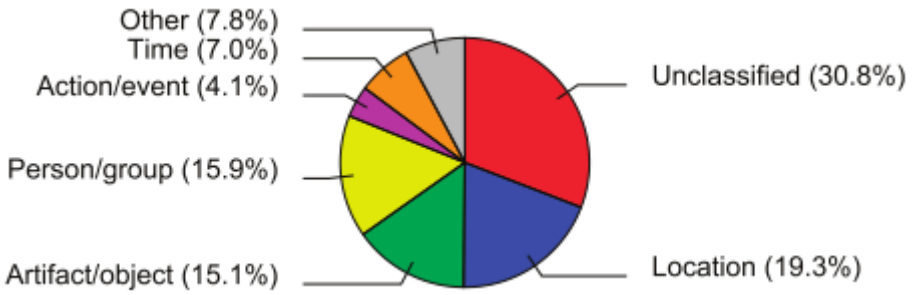
ข้อเสียของการทำการกรองร่วมคือ ไม่สามารถใช้กับระบบที่มีขนาดเล็กหรือระบบที่เพิ่งถูกสร้างเนื่องจากต้องทำการคำนวณจากข้อมูลปริมาณที่ค่อนข้างมาก

ระบบที่จะพัฒนาขึ้นจะใช้หลักการของการกรองร่วมอิงแท็ก (Tag-Based Collaborative Filtering) โดยจะใช้ข้อมูลของแท็กเป็นตัวแปรที่มีอยู่ในระบบประกอบกับพฤติกรรมการใช้บริการของผู้ใช้บริการมาใช้ในการคำนวณ เพื่อวิเคราะห์และคำนวณหาแหล่งข้อมูลที่ตรงกับความสนใจของผู้ใช้บริการ

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 Survey on Social Tagging Techniques [5]

M. Gupta et al. [5] ได้ทำการศึกษาพฤติกรรมการใช้แท็กในสังคมออนไลน์ในปัจจุบันและพบว่าแรงบันดาลใจในการระบุแท็กของผู้ใช้บริการในเครือข่ายสังคมออนไลน์นั้นมีความหลากหลายขึ้นอยู่กับจุดประสงค์และบริบทของแหล่งข้อมูลที่ต้องการระบุแท็ก เช่น การใช้สำหรับอ้างอิงในอนาคต การแบ่งปันข้อมูล การแสดงความเห็น หรือการจัดกลุ่ม ทั้งนี้ยังได้สรุปประเภทของแท็กที่พบในปัจจุบันว่ามีลักษณะในการระบุโดยขึ้นกับแหล่งข้อมูล บริบท หรือคุณลักษณะของแหล่งข้อมูลนั้นๆ ซึ่งจะสังเกตได้ว่าการที่ผู้ใช้บริการมีการระบุแท็กให้กับแหล่งข้อมูลที่เครือข่ายสังคมออนไลน์นั้นก็เปรียบเสมือนตัวแทนพฤติกรรมที่สอดคล้องกับความต้องการหรือความสนใจของผู้ใช้บริการ ณ ขณะเวลาหนึ่งๆ ดังภาพที่ 5



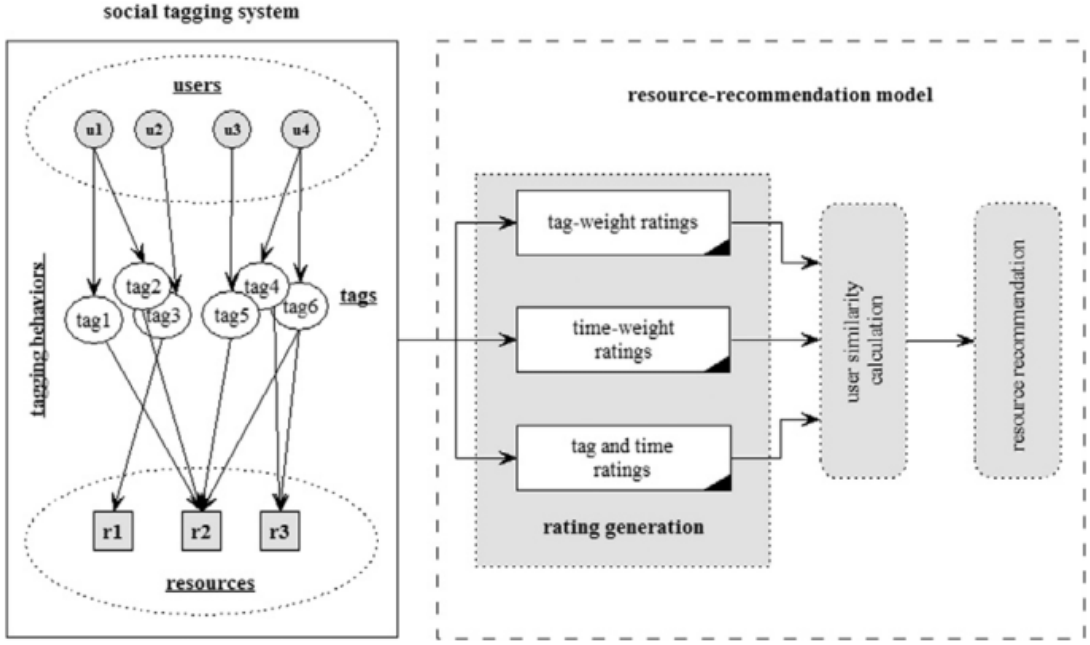
ภาพที่ 5 แผนภูมิแสดงสัดส่วนพฤติกรรมการใช้แท็กของผู้ใช้งานในสังคมออนไลน์ [5]

งานวิจัยดังกล่าวได้สรุปถึงแรงบันดาลใจและลักษณะของข้อมูลแท็กที่ถูกสร้างขึ้นในสังคมออนไลน์ ซึ่งเห็นได้ว่า ส่วนใหญ่แล้วข้อมูลแท็กนั้นจะสื่อถึงพฤติกรรม ความต้องการหรือความสนใจของผู้ใช้บริการในระบบ ทำให้ผู้วิจัยเล็งเห็นว่าจะเป็นการดีที่ระบบจะสามารถระบุแท็กที่ผู้ใช้บริการสนใจได้โดยอัตโนมัติ โดยดูจากพฤติกรรมการใช้บริการซึ่งจะสามารถเก็บความต้องการรองหรือความต้องการแฝงที่ผู้ใช้บริการไม่ได้ทำการระบุไว้ในระบบ

2.2.2 A Recommender System Based on Tag and Time Information for Social Tagging Systems [6]

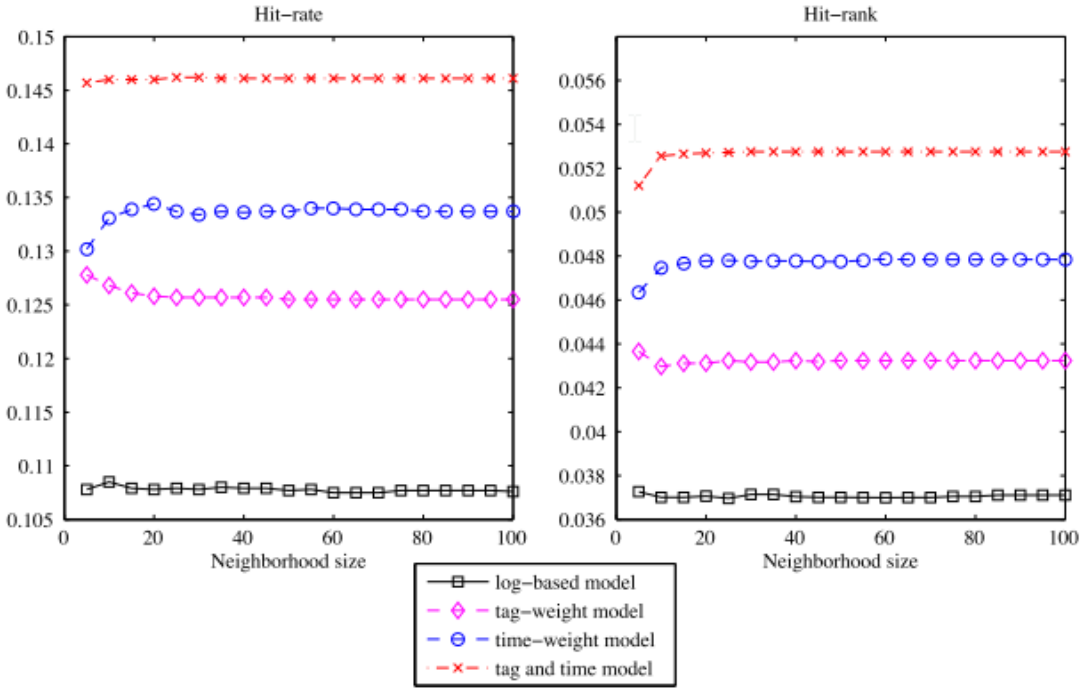
Zheng et al. [6] ได้นำเสนองานวิจัยของระบบแนะนำการสืบค้นงานวิจัยจากเว็บไซต์ CiteULike โดยใช้การวิเคราะห์ข้อมูลจากแท็กและเวลาบนเครือข่ายสังคมออนไลน์ โดยตั้งสมมติฐานว่า แท็กที่ผู้ใช้บริการเลือกเป็นตัวแทนของหัวข้อที่ผู้ใช้บริการให้ความสนใจ Zheng ได้นำเสนอวิธีการสร้างแท็กสเปซ (Tag Space) ของแต่ละผู้ใช้บริการเพื่อนำมาลดความซับซ้อนของการคำนวณคะแนน และสามารถนำมาใช้พิจารณาความสนใจ ณ ขณะเวลาปัจจุบันได้โดยง่าย เมื่อเทียบกับการวิเคราะห์แบบอื่นๆ ที่เน้นหนักไปทางการจัดกลุ่มของผู้ใช้บริการเพื่อหาความคล้ายคลึงกันและนำมาทำนายผลซึ่งต้องการข้อมูลในปริมาณมากและใช้ทรัพยากรค่อนข้างสูง

Zheng ได้ทำการทดลองโดยการสร้างระบบแนะนำผลโดยการเปรียบเทียบประสิทธิภาพในการแนะนำจากระบบที่วิเคราะห์และแนะนำจากค่าน้ำหนักของแท็ก ระบบที่วิเคราะห์และแนะนำจากค่าน้ำหนักของเวลา และระบบที่วิเคราะห์และแนะนำจากน้ำหนักของแท็กและเวลาร่วมกันดังภาพที่ 6



ภาพที่ 6 ภาพรวมของระบบแนะนำโดยใช้ข้อมูลแท็กและเวลา [6]

ผลการทดลองพบว่าระบบซึ่งวิเคราะห์และแนะนำจากค่าน้ำหนักแท็กและเวลาร่วมกันให้ความแม่นยำสูงที่สุดโดยให้ความแม่นยำสูงกว่า ระบบที่วิเคราะห์และแนะนำจากค่าน้ำหนักทางด้านเวลาหรือแท็กเพียงอย่างเดียวดังภาพที่ 7

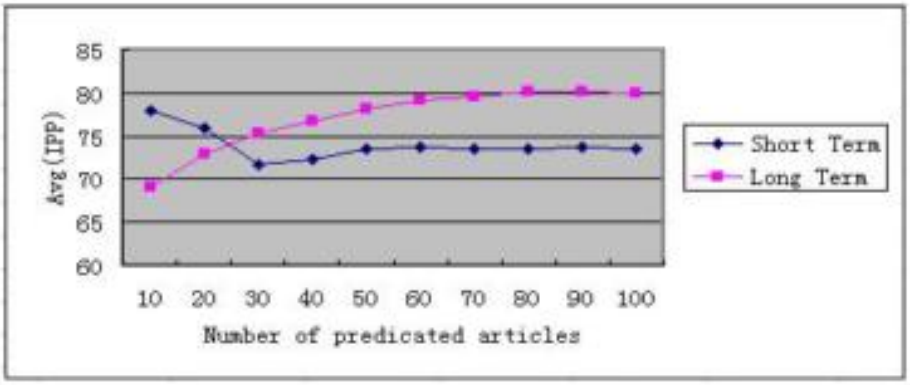


ภาพที่ 7 ค่าความแม่นยำจากระบบที่วิเคราะห์และแนะนำจากค่าน้ำหนักในรูปแบบต่างๆ [6]

ผู้วิจัยได้เล็งเห็นประโยชน์ของการสร้างแท็กสเปซซึ่งสามารถช่วยให้ระบบสามารถวิเคราะห์และคำนวณเพื่อการแนะนำแหล่งข้อมูลที่ผู้ใช้บริการสนใจ โดยในระบบที่จะพัฒนาขึ้นนั้นจะประยุกต์เอาการสร้างแท็กสเปซเป็นการสร้างแท็กคลาวด์เสมือนโดยที่แท็กคลาวด์เสมือนนั้นจะผูกติดอยู่กับผู้ใช้บริการและตัวแหล่งข้อมูล และจะมีการปรับค่าน้ำหนักของแท็กที่อยู่ในคลาวด์ตามพฤติกรรมของผู้ใช้บริการ

2.2.3 Model Bloggers' Interests Based on Forgetting Mechanism [7]

Cheng et al [7] ได้นำเสนอแบบจำลองในการวิเคราห์ความสนใจของผู้ใช้งานบล็อกโดยวิเคราะห์จากปัจจัยทางด้านเวลา Cheng ได้นำเสนอสมการการลืมเลือน (Forgetting Function) แบบระยะสั้น และระยะยาว (Short-Term และ Long-Term) โดยจากผลการทดลองพบว่าสมการการลืมเลือนแบบระยะสั้นจะเหมาะสมกับระบบที่ยังมีปริมาณข้อมูลน้อย ในทางกลับกันสมการการลืมเลือนแบบระยะยาวจะเหมาะสมกับระบบที่มีปริมาณข้อมูลค่อนข้างมากอย่างมีนัยสำคัญ ดังภาพที่ 8



ภาพที่ 8 แผนภูมิแสดงจำนวนบทความที่แนะนำตามขนาดของข้อมูล [7]

ระบบที่ได้นำเอาสมการการลืมเลือนแบบระยะสั้นมาประยุกต์ใช้ในการคำนวณค่าน้ำหนักของแท็กใหม่โดยสมการการลืมเลือนแบบระยะสั้นจะมีลักษณะดังสมการ (1)

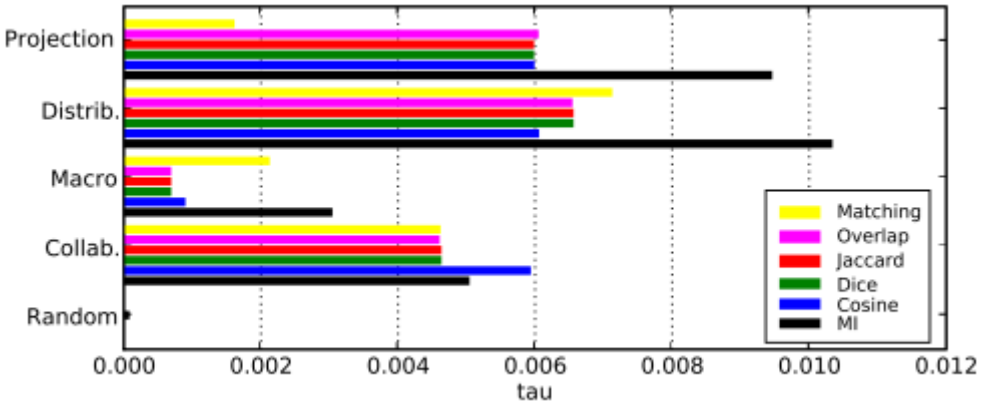
$$\mathbf{F(t)} = e^{-\frac{ln2 \times (t-est)}{hl}} \quad (1)$$

โดยที่ t คือวันที่ปัจจุบัน est คือวันที่มีการใช้งานครั้งสุดท้าย และ hl คือครึ่งชีวิต (half life) หรือระยะเวลาในการลืมเลือนซึ่งเป็นค่าคงที่ที่สามารถปรับเปลี่ยนได้ โดยที่ถ้าค่า hl มีค่ามากจะการลืมเลือนเป็นไปได้ช้าหรือค่าน้ำหนักที่ได้จากการคำนวณจะมีมากขึ้นนั่นเอง

2.2.4 Evaluating Similarity Measures for Emergent Semantics of Social Tagging [8]

B. Makines et al. [8] ได้ทำการศึกษาเพื่อทำการวิเคราะห์ประสิทธิภาพของสมการคำนวณความคล้ายกันระหว่างแหล่งข้อมูลโดยพิจารณาจากแท็กข้อมูลในรูปแบบต่างๆ เช่น สมการความคล้ายแบบแจ็กการ์ด (Jaccard) ไดซ์ (Dice) โคไซน์ (Cosine) และแบบสารสนเทศร่วม (Mutual Information) โดยพิจารณาจากเกณฑ์การรวบรวมข้อมูลในรูปแบบที่แตกต่างกัน เช่น แบบโปรเจคชัน (Projection) หรือการไม่นับข้อมูลซ้ำ แบบกระจาย (Distribution) และแบบร่วม (Collaborative) พบว่าสมการความคล้ายแบบโคไซน์ให้ผลลัพธ์ที่ดีที่สุดในการรวบรวมข้อมูลแบบร่วมดังภาพที่ 9

จากผลการทดลองในงานวิจัยนี้ ผู้วิจัยได้พิจารณานำเอาสมการคำนวณความคล้ายแบบโคไซน์มาใช้คำนวณความคล้ายกันระหว่างผู้ใช้บริการกับผู้ใช้บริการ และผู้ใช้บริการกับแหล่งข้อมูลเพื่อจะเอาไปเป็นส่วนหนึ่งในการคำนวณเพื่อที่จะแนะนำแหล่งข้อมูลให้กับผู้ใช้บริการต่อไป



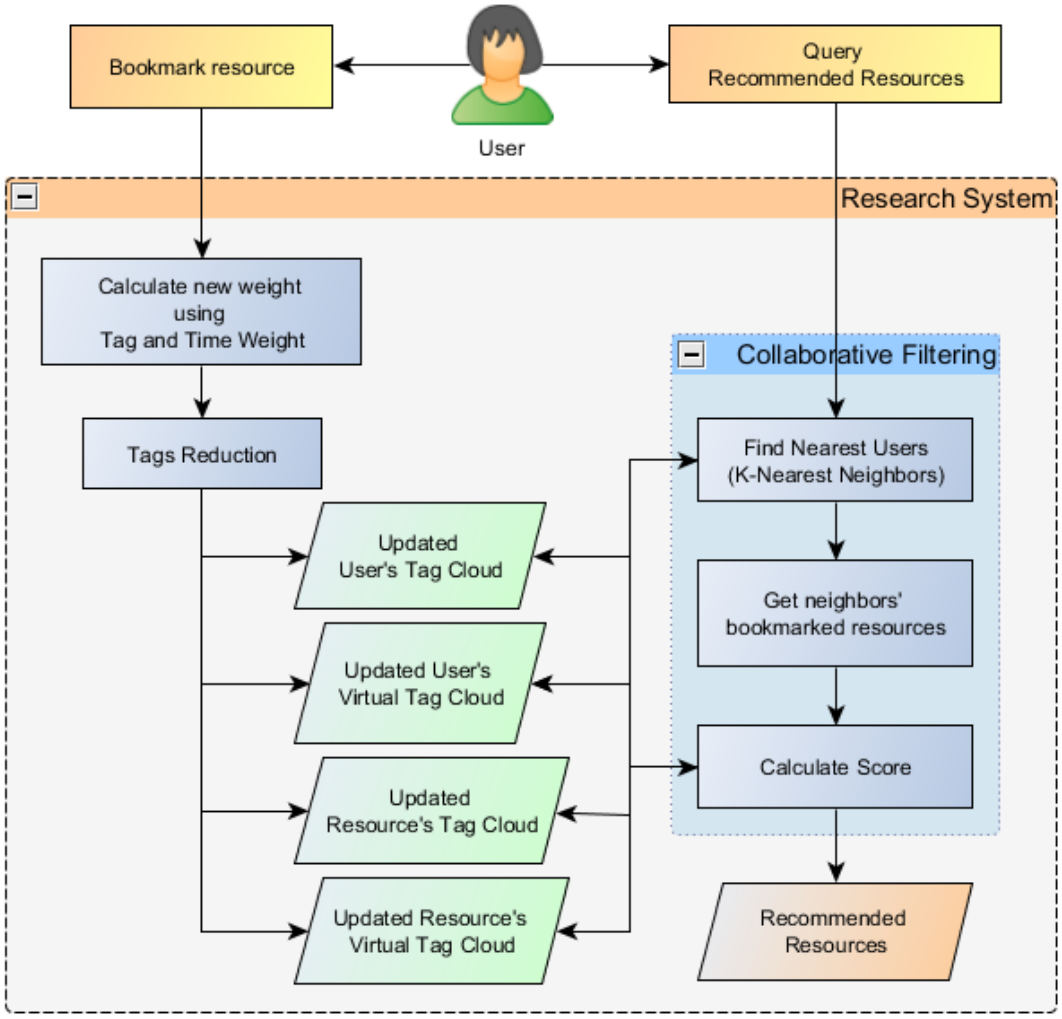
ภาพที่ 9 แผนภูมิแสดงค่าความแม่นยำของสมการความคล้ายในสภาพแวดล้อมของระบบที่ทำการทดลอง [8]

บทที่ 3
วิธีดำเนินการวิจัย

3.1 แนวคิดในการพัฒนา

ระบบที่จะพัฒนาขึ้นได้นำเสนอวิธีการสร้างแท็กคลาวด์เสมือน เพื่อใช้วิเคราะห์ความต้องการรองหรือความต้องการแฝง และใช้คำนวณร่วมกับแท็กข้อมูลที่มีอยู่จริง เพื่อเพิ่มความแม่นยำในการแนะนำแหล่งข้อมูลให้ตรงกับความสนใจหรือความต้องการของผู้ใช้บริการ โดยระบบจะทำการสร้างแท็กคลาวด์เสมือนให้กับผู้ใช้บริการและแหล่งข้อมูล ซึ่งจะมีการปรับเปลี่ยนข้อมูลแท็กตามพฤติกรรมการใช้งานของผู้ใช้บริการในระบบ

ภาพรวมการทำงานจะแบ่งออกเป็นสองส่วนคือกรณีผู้ใช้งานเข้ามาแท็กแหล่งข้อมูลในระบบและผู้ใช้งานต้องการเลือกชมรายการแหล่งข้อมูลที่ระบบแนะนำดังภาพที่ 10



ภาพที่ 10 ภาพรวมการทำงานของระบบ

กรณีแรก: ผู้ใช้งานเข้ามาแท็กแหล่งข้อมูลในระบบ ประกอบด้วยขั้นตอน ดังนี้

1. ระบบทำการปรับเปลี่ยนน้ำหนักของแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือน โดยอิงจากพฤติกรรมการแท็กและปัจจัยทางด้านเวลา (วิธีการคำนวณค่าน้ำหนักของแท็กอธิบายในหัวข้อ 3.2)
2. ระบบทำการตัดทอนแท็กที่มีค่าน้ำหนักน้อยกว่าค่าขีดแบ่งออกจากแท็กคลาวด์และแท็กคลาวด์เสมือนเพื่อตัดแท็กที่ไม่มีส่วนเกี่ยวข้องออก และควบคุมปริมาณแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนไม่ให้มีมากเกินไป
3. ระบบทำการจัดเก็บข้อมูลแท็กคลาวด์และแท็กคลาวด์เสมือนของแหล่งข้อมูลและผู้ใช้บริการที่ได้จากการปรับค่าน้ำหนักและการตัดทอนลงฐานข้อมูล

กรณีที่ 2: ผู้ใช้งานต้องการเลือกชมรายการแหล่งข้อมูลที่ระบบแนะนำ

ระบบจะใช้วิธีการกรองร่วมเพื่อกรองข้อมูลเฉพาะกลุ่มผู้ใช้บริการและแหล่งข้อมูลที่สนใจ โดยประกอบด้วยขั้นตอนดังนี้

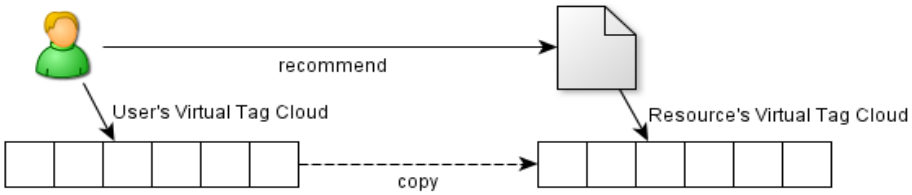
1. ระบบทำการค้นหาผู้ใช้งานอื่นๆ ที่มีแท็กคลาวด์หรือแท็กคลาวด์เสมือนคล้ายกับผู้ใช้งาน
2. ระบบทำการค้นหาแหล่งข้อมูลที่ผู้ใช้งานที่มีความคล้ายกับผู้ใช้งานเคยทำรายการแท็กมาก่อน
3. ระบบทำการคำนวณความคล้ายระหว่างผู้ใช้งานและรายการแหล่งข้อมูลที่ได้จากข้อ 2 โดยจัดเรียงแหล่งข้อมูลตามค่าความคล้ายที่คำนวณได้จากมากไปน้อยและแสดงผลให้กับผู้ใช้งาน (วิธีการคำนวณค่าความคล้ายอธิบายในหัวข้อ 3.3)

3.2 การจัดการแท็กคลาวด์เสมือนและการปรับน้ำหนัก

นอกจากการจัดการข้อมูลแท็กคลาวด์จริงแล้ว ระบบที่พัฒนาจะทำการสร้างแท็กคลาวด์เสมือนขึ้นโดยดูพิจารณาจากพฤติกรรมการใช้บริการของผู้ใช้บริการ หลักการในการสร้างและปรับปรุงแท็กในคลาวด์เสมือนมีดังนี้

3.2.1 การแนะนำแหล่งข้อมูลโดยผู้ใช้บริการ

การแนะนำแหล่งข้อมูลโดยผู้ใช้บริการนั้น ระบบจะทำการติดแท็กให้กับคลาวด์เสมือนของแหล่งข้อมูลหรือข้อมูลนั้นโดยอัตโนมัติ โดยแท็กข้อมูลที่จะทำการติดนั้นจะถูกนำมาจากแท็กข้อมูลทั้งหมดที่ติดกับตัวของผู้ใช้บริการทั้งหมด ด้วยสมมติฐานที่ว่าสิ่งที่ผู้ใช้บริการแนะนำนั้นคือสิ่งที่ผู้ใช้บริการให้ความสนใจอยู่ ดังภาพที่ 11

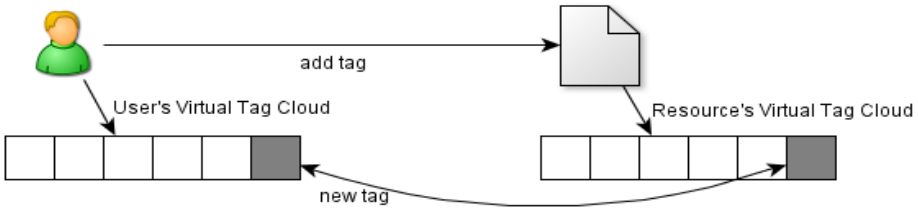


ภาพที่ 11 การสร้างแท็กคลาวด์เสมือนของแหล่งข้อมูลเมื่อมีการเพิ่มแหล่งข้อมูลเข้าสู่ระบบ

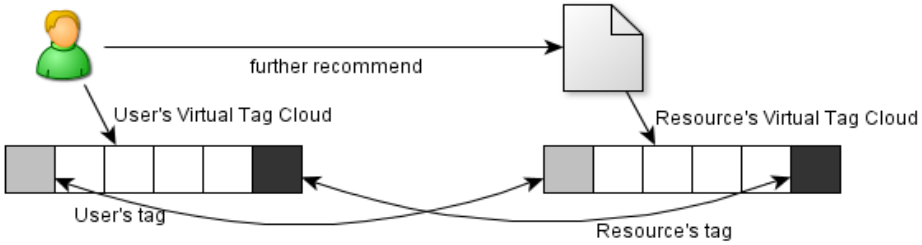
3.2.2 การเพิ่มแท็กข้อมูล

การปรับเพิ่มแท็กข้อมูลของผู้ใช้บริการจะกระทำเมื่อผู้ใช้บริการมีการเพิ่มแท็กให้กับแหล่งข้อมูลหรือทำการแนะนำแหล่งข้อมูลต่อ ในกรณีของการเพิ่มแท็กให้กับแหล่งข้อมูลนั้น ระบบจะทำการเพิ่มแท็กข้อมูลไปยังแท็กคลาวด์เสมือนของแหล่งข้อมูลและหากยังไม่พบข้อมูลแท็กดังกล่าวในแท็กคลาวด์เสมือนของผู้ใช้บริการ ระบบจะทำการเพิ่มข้อมูลแท็กนั้นไปยังแท็กคลาวด์ของผู้ใช้บริการด้วยดังภาพที่ 12

ในกรณีของการเพิ่มข้อมูลแท็กจากการแนะนำต่อ ระบบจะทำการตรวจสอบข้อมูลแท็กในแท็กคลาวด์เสมือนของทั้งผู้ใช้บริการและแหล่งข้อมูลโดยจะทำการเพิ่มข้อมูลแท็กที่ติดกับแหล่งข้อมูลที่ไม่อยู่ในแท็กคลาวด์เสมือนของผู้ใช้บริการเข้าไปในแท็กคลาวด์เสมือนของผู้ใช้บริการ ในทางกลับกันก็จะเพิ่มข้อมูลแท็กที่อยู่ในแท็กคลาวด์เสมือนของผู้ใช้บริการแต่ไม่อยู่ในแท็กคลาวด์เสมือนของแหล่งข้อมูลเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล ดังภาพที่ 13



ภาพที่ 12 เพิ่มแท็กในแท็กคลาวด์เสมือนเมื่อมีการเพิ่มแท็กให้กับแหล่งข้อมูล



ภาพที่ 13 การเพิ่มแท็กในแท็กคลาวด์เสมือนเมื่อมีการแนะนำต่อ

3.2.3 การปรับเปลี่ยนน้ำหนักของแท็กในแท็กคลาวด์เสมือน

การปรับเปลี่ยนน้ำหนักของแท็กในคลาวด์เสมือนจะถูกกระทำก็ต่อเมื่อมีการใช้งานแท็กจากเหตุการณ์ในข้อ 3.1.1 และ 3.1.2 ข้างต้น โดยการค่าน้ำหนักจะประกอบด้วยค่าน้ำหนักจากการคำนวณน้ำหนักจากการใช้งานแท็กและค่าน้ำหนักจากเวลา (Tag Weight และ Time Weight) สมการ (2) แสดงวิธีคำนวณค่าน้ำหนักใหม่ของแท็กแต่ละตัวจากการใช้งานแท็ก

$$\textbf{ค่าน้ำหนักจากการใช้งานแท็ก} = \frac{\textbf{จำนวนครั้งที่แท็กได้ถูกใช้งาน}}{\textbf{จำนวนการใช้งานแท็กทั้งหมดในแท็กคลาวด์}} \quad (2)$$

สมการ (3) แสดงวิธีการคำนวณค่าน้ำหนักของแท็กอิงปัจจัยทางด้านเวลาโดยใช้สมการการลืมเลือนจากหัวข้อ 2.2.3

$$\textbf{ค่าน้ำหนักจากเวลา} = e^{-\frac{ln2 \times (t-est)}{hl}} \quad (3)$$

โดยที่ *t* คือวันที่ปัจจุบัน *est* คือวันที่แท็กถูกใช้งานครั้งสุดท้าย และ *hl* คือครึ่งชีวิต (Half Life) หรือระยะเวลาในการลืมเลือน (หน่วยเป็นวัน)

สมการ (4) แสดงวิธีการคำนวณน้ำหนักใหม่ของแท็กโดยใช้ทั้งปัจจัยเรื่องการใช้งานแท็กและปัจจัยทางด้านเวลา

$$\textbf{ค่าน้ำหนักใหม่ของแท็ก} = \textbf{ค่าน้ำหนักจากการใช้งานแท็ก} \times \textbf{ค่าน้ำหนักจากเวลา} \quad (4)$$

หากค่าน้ำหนักใหม่ที่ได้มีค่าน้อยกว่าค่าขีดแบ่ง ระบบจะทำการลบข้อมูลแท็กนั้นออกจากแท็กคลาวด์ เนื่องจากมีน้ำหนักน้อยจนไม่จำเป็นต้องนำมาใช้ร่วมในการคำนวณ การทดลองจะมีการปรับค่าขีดแบ่งให้มีความหลากหลายเพื่อทำการทดสอบหาค่าที่เหมาะสม

เมื่อมีการใช้งานระบบไปในช่วงเวลาหนึ่ง การปรับน้ำหนักและการตัดทอนข้อมูลแท็กที่ไม่จำเป็นจะทำให้จำนวนแท็กที่อยู่ในแท็กคลาวด์เสมือนเข้าใกล้จำนวนที่เหมาะสม

ระบบที่พัฒนาขึ้นได้ทำการทดลองปรับค่าน้ำหนัก 2 วิธี คือ จากข้อมูลแท็กคลาวด์จริงและข้อมูลแท็กคลาวด์เสมือน เพื่อใช้ในการเปรียบเทียบประสิทธิภาพของการวิเคราะห์และคำนวณผลการแนะนำแหล่งข้อมูลที่ได้

3.3 การแนะนำแหล่งข้อมูล

การแนะนำแหล่งข้อมูลจะใช้วิธีการกรองร่วมเพื่อเลือกเฉพาะข้อมูลที่สนใจเพื่อลดปริมาณข้อมูลที่จะต้องนำมาพิจารณา ผู้วิจัยมีแนวคิดที่จะใช้วิธีการหาเพื่อนบ้านจำนวน K ตัวที่ใกล้ที่สุด (K-Nearest Neighbors) ระหว่างผู้ใช้บริการกับผู้ใช้บริการโดยการคำนวณความคล้ายระหว่างแท็กคลาวด์และแท็กคลาวด์เสมือน การคำนวณความคล้ายจะใช้ผลลัพธ์ของการคำนวณค่า

น้ำหนักที่ได้จากการคำนวณในสมการ (4) มาคำนวณความคล้ายแบบโคไซน์โดยใช้ค่าแท็กข้อมูลและค่าน้ำหนักของแท็กที่มีอยู่ในแท็กคลาวด์ร่วมกัน

3.3.1 การคำนวณความคล้ายระหว่างผู้ใช้บริการกับผู้ใช้บริการ

งานวิจัยนี้ใช้สมการคำนวณความคล้ายโคไซน์ ซึ่งมีสูตรการคำนวณดังแสดงในสมการ (5) เพื่อคำนวณความคล้ายกันระหว่างผู้ใช้บริการกับผู้ใช้บริการ

$$\mathbf{similarity}\ (\mathbf{u},\mathbf{v}) = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{v}}}{\|\vec{\mathbf{u}}\|\cdot\|\vec{\mathbf{v}}\|} = \frac{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{v})} \mathbf{W}_{\mathbf{u},\mathbf{x}} \times \mathbf{W}_{\mathbf{v},\mathbf{x}}}{\sqrt{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{v})} \mathbf{W}_{\mathbf{u},\mathbf{x}}^{\mathbf{2}}} \times \sqrt{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{v})} \mathbf{W}_{\mathbf{v},\mathbf{x}}^{\mathbf{2}}}} \qquad (5)$$

โดยที่ X(u,v) คือเซตของผู้ใช้บริการ v ที่มีข้อมูลในแท็กคลาวด์ร่วมกันกับผู้ใช้บริการ u

หากค่าความคล้ายเท่ากับ -1 แสดงว่า แท็กคลาวด์มีความแตกต่างกันอย่างชัดเจน หากเท่ากับ 1 แสดงว่า แท็กคลาวด์มีความเหมือนกันทุกประการ และหากเท่ากับ 0 แสดงว่าข้อมูลในแท็กคลาวด์เป็นอิสระต่อกัน

3.3.2 การคำนวณความคล้ายระหว่างผู้ใช้บริการและแหล่งข้อมูล

เมื่อได้จำนวนเพื่อนบ้าน (ผู้ใช้งาน) ที่มีความคล้ายจำนวน K คนแล้วระบบจะทำการค้นหาแหล่งข้อมูลที่ผู้ใช้งานเหล่านั้นเคยทำการแท็ก เพื่อนำมาวิเคราะห์ความคล้ายและแนะนำให้กับผู้ใช้งานเป้าหมาย โดยจะคำนวณความคล้ายระหว่างผู้ใช้งานกับแหล่งข้อมูลโดยใช้สมการโคไซน์ดังสมการ (6)

$$\mathbf{similarity}\ (\mathbf{u},\mathbf{r}) = \frac{\vec{\mathbf{u}}\cdot\vec{\mathbf{r}}}{\|\vec{\mathbf{u}}\|\cdot\|\vec{\mathbf{r}}\|} = \frac{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{r})} \mathbf{W}_{\mathbf{u},\mathbf{x}} \times \mathbf{W}_{\mathbf{r},\mathbf{x}}}{\sqrt{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{r})} \mathbf{W}_{\mathbf{u},\mathbf{x}}^{\mathbf{2}}} \times \sqrt{\sum_{\mathbf{x}\in\mathbf{X}(\mathbf{u},\mathbf{r})} \mathbf{W}_{\mathbf{r},\mathbf{x}}^{\mathbf{2}}}} \qquad (6)$$

โดยที่ X(u,r) คือเซตของแหล่งข้อมูล r ที่มีข้อมูลในแท็กคลาวด์ร่วมกันกับผู้ใช้บริการ u

กำหนดให้ score(u,r) คือ คะแนนของแหล่งข้อมูล ซึ่งมีค่าเท่ากับค่าความคล้ายระหว่างผู้ใช้งานและแหล่งข้อมูล ดังสมการ (7)

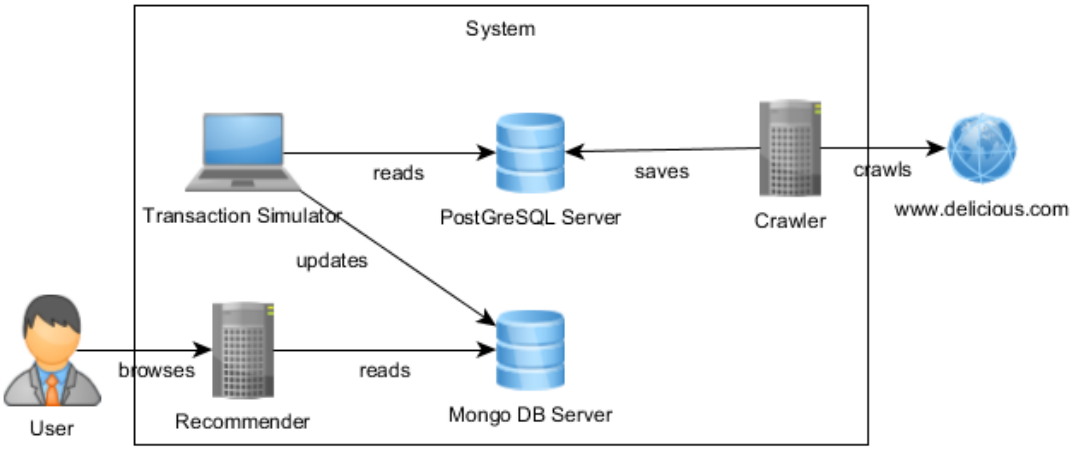
$$\boldsymbol{score}(\boldsymbol{u},\boldsymbol{r}) = \boldsymbol{similarity}(\boldsymbol{u},\boldsymbol{r}) \qquad (7)$$

การแนะนำผลจะแนะนำแหล่งข้อมูลที่มีคะแนน score (u,r) สูงที่สุดขึ้นมาก่อน โดยเรียงลำดับจากมากไปหาน้อย

บทที่ 4
การออกแบบและพัฒนาระบบ

4.1 สถาปัตยกรรมระบบ

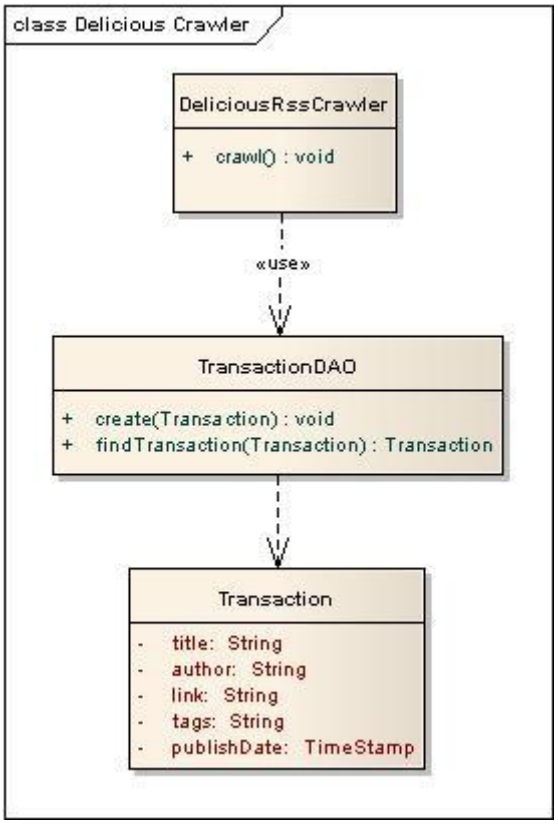
ระบบที่พัฒนาขึ้นแบ่งการทำงานออกเป็น 3 ส่วนคือ 1) ส่วนการดึงข้อมูลจากเว็บไซต์ Delicious เพื่อนำมาใช้ในการทดสอบสมมติฐาน 2) ส่วนการจำลองการใช้งานเว็บไซต์ Delicious เพื่อสร้างแท็กคลาวด์และแท็กคลาวด์เสมือน 3) ส่วนการแนะนำแหล่งข้อมูลและแสดงผลในรูปแบบของเว็บแอปพลิเคชันดังภาพที่ 14



ภาพที่ 14 สถาปัตยกรรมระบบ

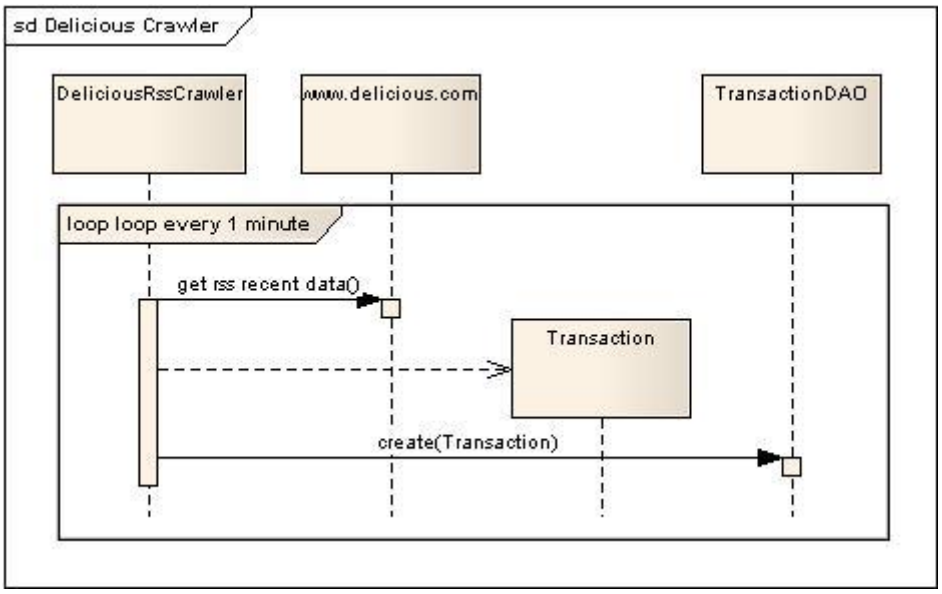
4.1.1 ส่วนการดึงข้อมูลจากเว็บไซต์ Delicious.com

ในส่วนนี้จะเป็นส่วนที่เชื่อมต่อกับเว็บไซต์ Delicious.com เพื่อทำการดึงข้อมูลการใช้บริการล่าสุด โดยจะทำการวนซ้ำเป็นรอบๆ ทุกๆ 1 นาทีเพื่อทำการเก็บข้อมูลจากเว็บไซต์ Delicious.com ที่ URL http://feeds.delicious.com/v2/rss/recent?min=2&count=500 โดยหลังจากที่ดึงข้อมูลมาแล้วจะทำการเก็บข้อมูลลงยังฐานข้อมูล PostGreSQL Server โดยโครงสร้างหลักของส่วนที่ดึงข้อมูลจะเป็นไปตามคลาสไดอะแกรมดังภาพที่ 15



ภาพที่ 15 คลาสไดอะแกรมของส่วนที่ดึงข้อมูลจากเว็บไซต์ Delicious.com

การทำงานของส่วนนี้จะเป็นไปตามแผนภาพลำดับเหตุการณ์ดังภาพที่ 16 โดยจะทำซ้ำทุกๆ 1 นาที

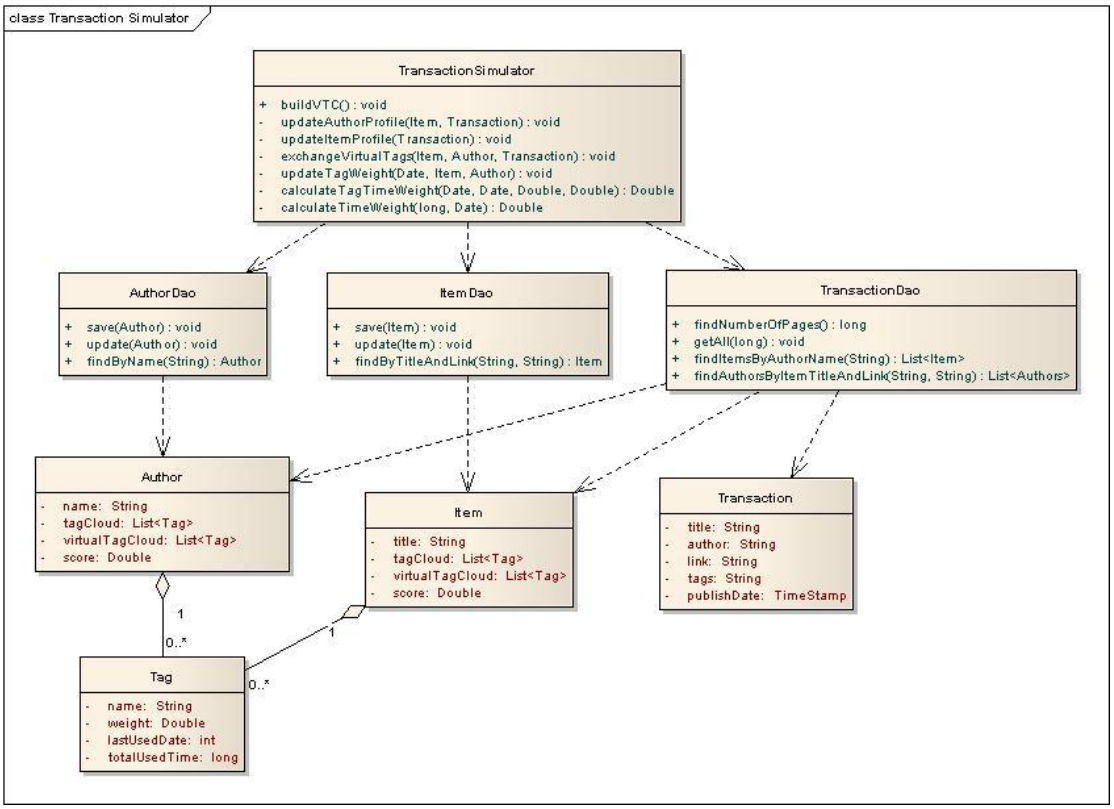


ภาพที่ 16 แผนภาพลำดับเหตุการณ์ของส่วนดึงข้อมูลจากเว็บไซต์ Delicious.com

4.1.2 ส่วนการจำลองการใช้งานของเว็บไซต์ Delicious.com

ในส่วนนี้เป็นส่วนของการจำลองการใช้งานของผู้ใช้บริการเว็บไซต์ Delicious.com เพื่อทำการสร้างแท็กคลาวด์และแท็กคลาวด์เสมือนของผู้ใช้บริการและแหล่งข้อมูล โดยจะรวมไปถึงกระบวนการแลกเปลี่ยนและปรับค่าน้ำหนักของแท็กของผู้ใช้งานและแหล่งข้อมูลตามพฤติกรรมการใช้งานและปัจจัยทางด้านเวลา

โครงสร้างหลักของการทำงานส่วนนี้จะเป็นไปตามคลาสไดอะแกรมดังภาพที่ 17 และการทำงานของส่วนนี้จะเป็นไปตามแผนภาพลำดับเหตุการณ์ดังภาพที่ 18



ภาพที่ 17 คลาสไดอะแกรมของส่วนที่จำลองการใช้งานของผู้ใช้งานเว็บไซต์ Delicious.com

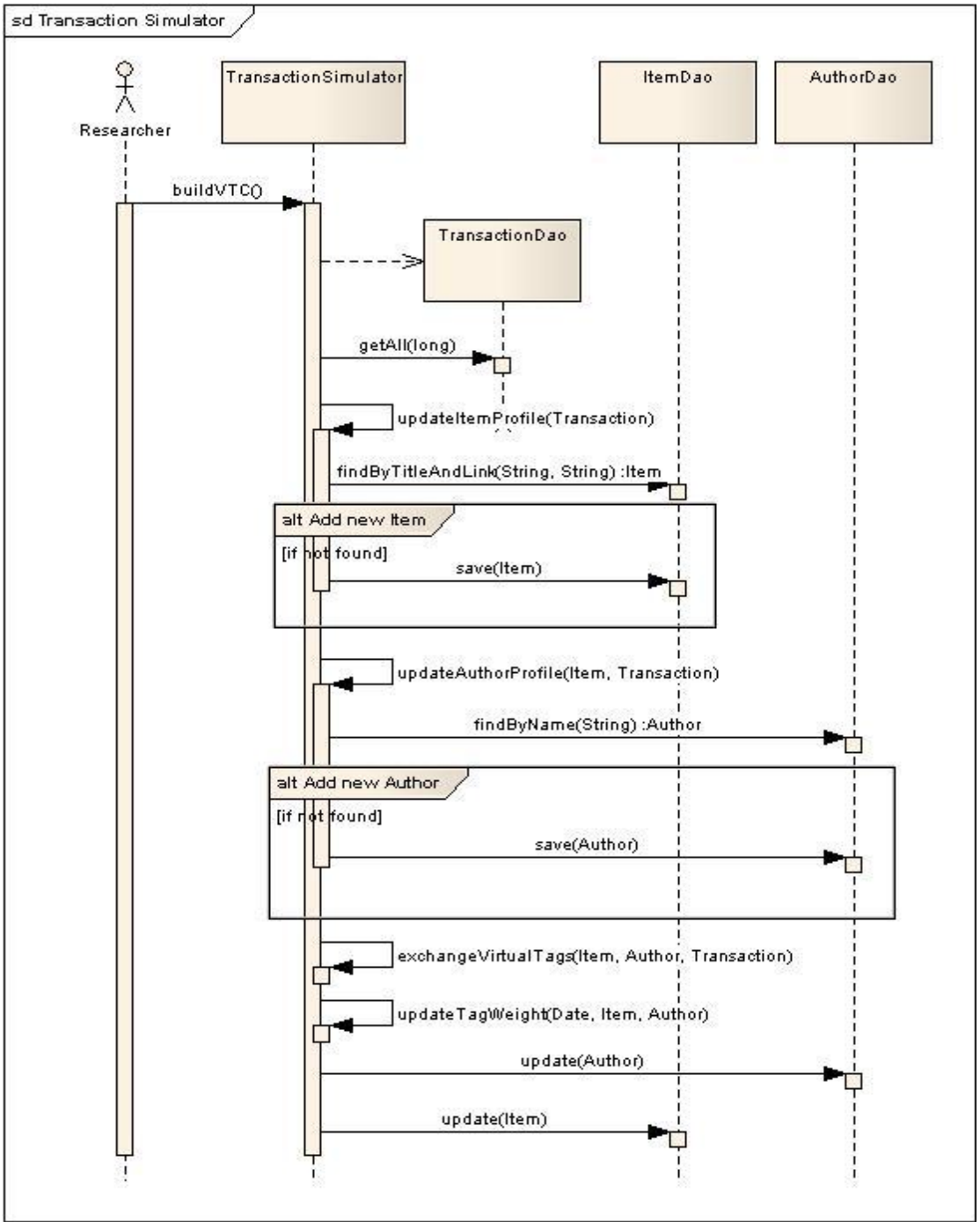
4.1.3 ส่วนการแนะนำแหล่งข้อมูลและแสดงผล

ในส่วนนี้เป็นส่วนที่คำนวณข้อมูลที่มีในระบบเพื่อใช้ในการแนะนำแหล่งข้อมูลให้กับผู้ใช้บริการ โดยจะประกอบด้วยส่วนของการแนะนำแหล่งข้อมูลและส่วนเชื่อมต่อผู้ใช้บริการ

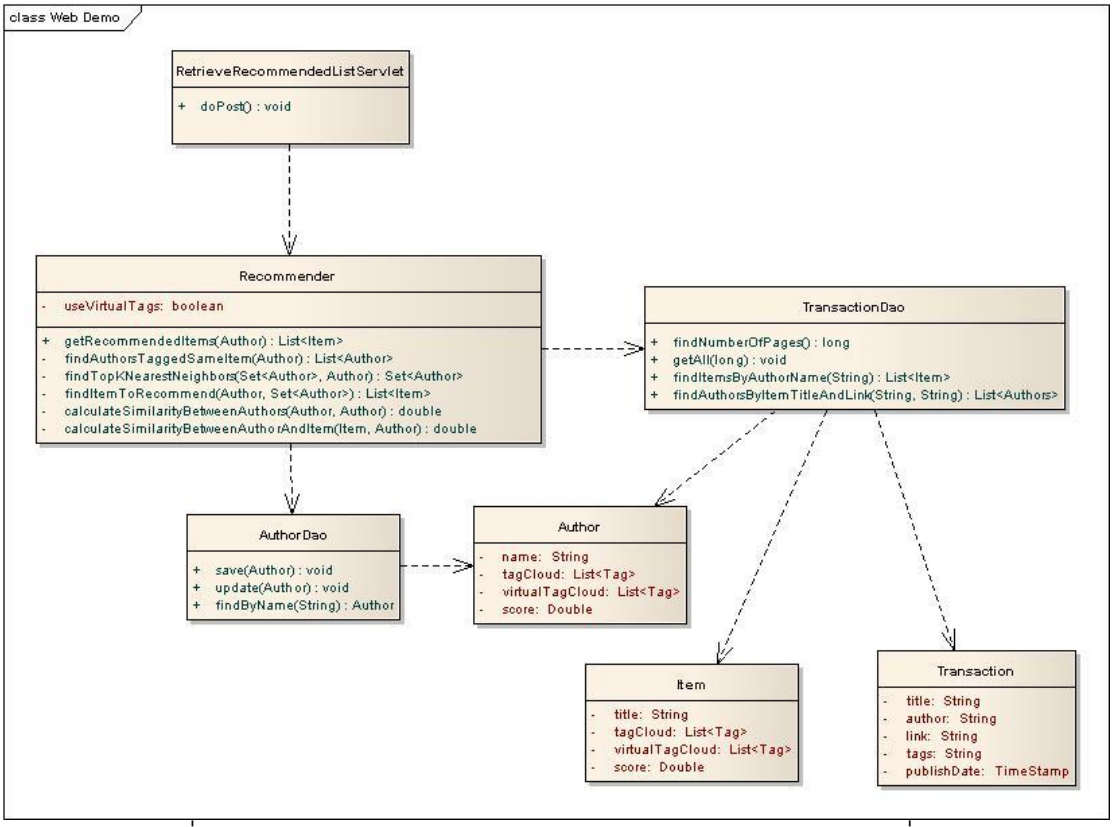
ในส่วนของการแนะนำแหล่งข้อมูลได้ใช้วิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (Top K-Nearest Neighbors) โดยการคำนวณความคล้ายกันระหว่างผู้ใช้บริการและทำการค้นหาแหล่งข้อมูลที่ผู้ใช้บริการที่มีความคล้ายกับผู้ใช้บริการเป้าหมายเคยทำการแท็กเพื่อนำมาคำนวณความคล้ายระหว่างผู้ใช้บริการกับแหล่งข้อมูลและนำค่าที่ได้หลังจากการคำนวณไปวิเคราะห์เพื่อแสดงผลการแนะนำต่อไป

ในส่วนการแสดงผลให้ผู้ใช้บริการมีการแสดงผลในรูปแบบของเว็บแอปพลิเคชัน โดยใช้ Java Servlet และ JSP เป็นเครื่องมือในการพัฒนา

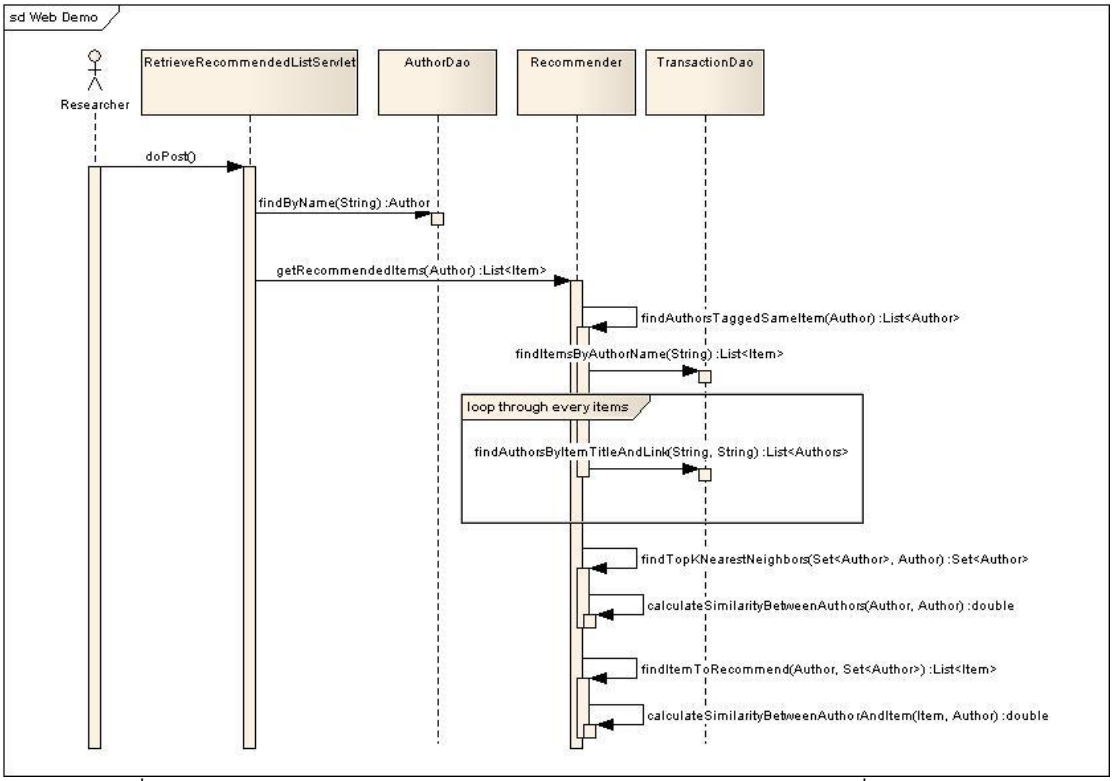
โครงสร้างหลักของการทำงานส่วนนี้จะเป็นไปตามคลาสไดอะแกรมดังภาพที่ 19 และการทำงานของส่วนดังกล่าวจะเป็นไปตามแผนภาพลำดับเหตุการณ์ดังภาพที่ 20



ภาพที่ 18 แผนภาพลำดับเหตุการณ์ของส่วนที่จำลองการใช้งานของผู้ใช้งานเว็บไซต์ Delicious.com



ภาพที่ 19 คลาสไดอะแกรมของส่วนของการแนะนำแหล่งข้อมูลและเชื่อมต่อผู้ใช้งาน



ภาพที่ 20 แผนภาพลำดับเหตุการณ์ของส่วนแนะนำแหล่งข้อมูลและเชื่อมต่อผู้ใช้งาน

4.2 สภาพแวดล้อมและเครื่องมือที่ใช้ในการพัฒนา

สภาพแวดล้อมและเครื่องมือที่ใช้ในการพัฒาระบบประกอบด้วยรายการฮาร์ดแวร์และซอฟต์แวร์ดังต่อไปนี้

ฮาร์ดแวร์

1. หน่วยประมวลผลอินเทล คอร์ 2 ดูโอ 2.40 กิกะเฮิร์ต (CPU Intel Core 2 Duo 2.40GHz)
2. หน่วยความจำ 4 กิกะไบต์ (4 GB RAM)
3. ฮาร์ดดิสก์แบบโซลิดสเตทความจุ 120 กิกะไบต์ (120 GB Solid-State Drive)

ซอฟต์แวร์

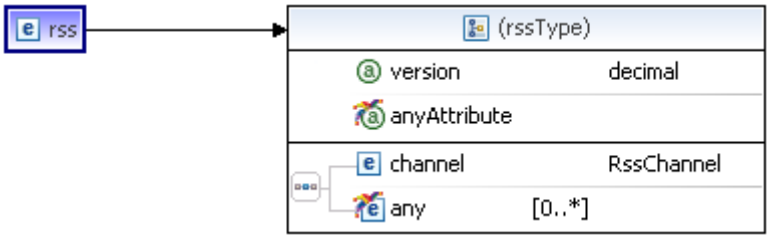
1. ระบบปฏิบัติการลีนุกซ์อูบันตู 10.04 (Ubuntu 10.04) แบบ 64 บิท
2. เครื่องมือพัฒนาโปรแกรมอีคลิปส์ 3.7.1 (Eclipse 3.7.1)
3. โปรแกรมชุดเครื่องมือพัฒนาจาวา 7 (Java Development Kit 7)
4. ระบบจัดการฐานข้อมูล PostGreSQL Server 8.4
5. ระบบจัดการฐานข้อมูล Mongo DB 2.0.2
6. เครื่องมือพัฒนาฐานข้อมูลเชิงวัตถุ Hibernate 3.0
7. โปรแกรมเว็บเซิร์ฟเวอร์ Apache Tomcat 7.0.23
8. เครื่องมือพัฒนาจาวาสคริปต์ YUI 2.9.0 (JavaScript YUI Library)
9. เครื่องมือการสร้างแผนภูมิด้วยภาษาจาวา JFreeChart

4.3 การพัฒนาระบบ

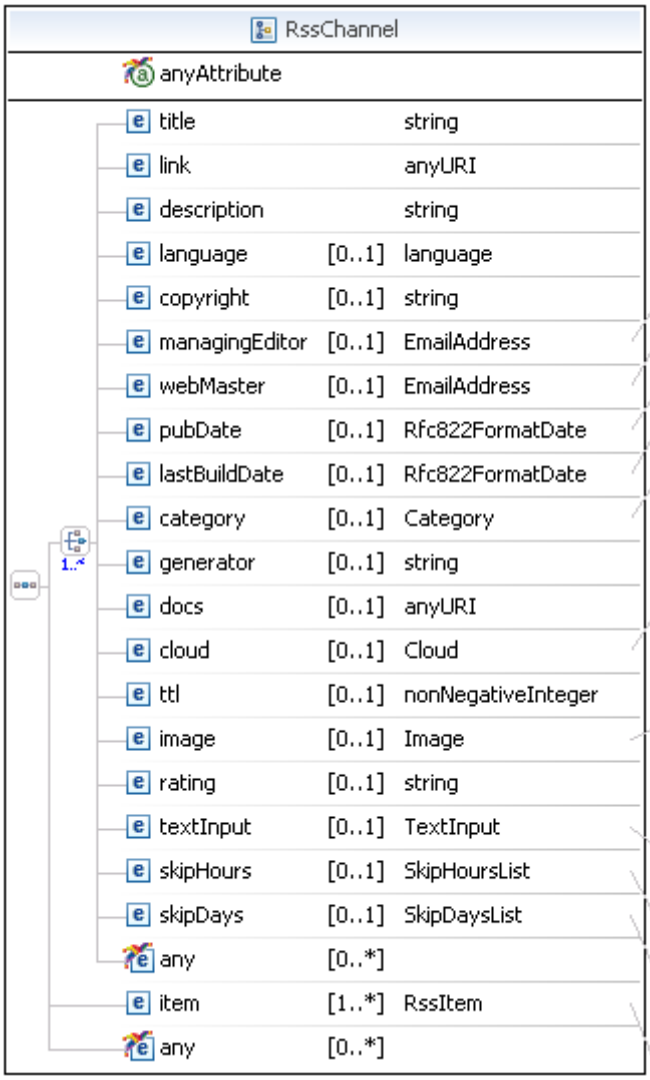
การพัฒนาระบบแบ่งออกเป็น 3 ส่วนหลักๆ ตามโครงสร้างของสถาปัตยกรรมระบบที่กล่าวในหัวข้อ 4.2 คือ 1) ส่วนการดึงข้อมูลจากเว็บไซต์ Delicious.com 2) ส่วนการจำลองการใช้งานเว็บไซต์ Delicious.com และ 3) ส่วนการแนะนำแหล่งข้อมูลและแสดงผล

4.3.1 ส่วนการดึงข้อมูลจากเว็บไซต์ Delicious.com

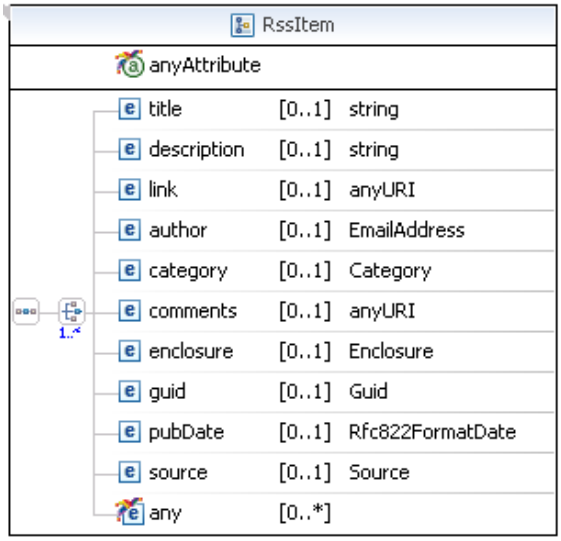
ดังที่กล่าวไว้ข้างต้นในหัวข้อ 4.1 การทำงานหลักของส่วนนี้คือการดึงข้อมูลจากเว็บไซต์ Delicious.com ด้วยโปรโตคอล HTTP ผ่านช่องทาง RSS (Really Simple Syndication) โดยลักษณะข้อมูลที่ได้รับจะอยู่ในรูปแบบของภาษา XML ซึ่งมีโครงสร้างดังภาพที่ 21, 22 และ 23 โดยมีโหนดที่สนใจคือ 1) rss ซึ่งเป็นโหนดที่อยู่บนสุด 2) โหนด channel เป็นโหนดรองลงมา และ 3) โหนด item ซึ่งเป็นโหนดที่ระบุรายละเอียดเกี่ยวกับแหล่งข้อมูล



ภาพที่ 21 โครงสร้างของข้อมูล XML ที่โหนด rss



ภาพที่ 22 โครงสร้างของข้อมูล XML ที่โหนด channel



ภาพที่ 23 โครงสร้างของข้อมูล XML ที่โหนด Item

ตัวอย่างของข้อมูล XML ที่ได้รับจากการร้องขอไปยังช่องทาง RSS ของเว็บไซต์ Delicious.com มีลักษณะดังภาพที่ 24

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0" xmlns:atom="http://www.w3.org/2005/Atom"
     xmlns:content="http://purl.org/rss/1.0/modules/content/"
     xmlns:wfw="http://wellformedweb.org/CommentAPI/"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:dc="http://purl.org/dc/elements/1.1/"
     xmlns:cc="http://web.resource.org/cc/">
    <channel>
        <title>Delicious recent</title>
        <link>http://www.delicious.com/recent</link>
        <description>the most recent bookmarks</description>
        <atom:link rel="self" type="application/rss+xml" href="http://feeds.delicious.com/v2/rss/recent?min=2&amp;count=500"/>

        <item>
            <title>Interview with Silvia Federici on Feminism, Reproduction and the Occupy Movement « Max Haiven</title>
            <pubDate>Tue, 17 Jan 2012 14:23:20 +0000</pubDate>
            <guid isPermaLink="false">http://www.delicious.com/url/0a04339ed7d494ba19b32f1d6f8b9ee4#vreteno</guid>
            <link>http://maxhaiven.com/2011/11/26/interview-with-silvia-federici-on-feminism-reproduction-and-the-occupy-movement/</link>
            <dc:creator><![CDATA[vreteno]]></dc:creator>
            <comments>http://www.delicious.com/url/0a04339ed7d494ba19b32f1d6f8b9ee4</comments>
            <wfw:commentRss>http://feeds.delicious.com/v2/rss/url/0a04339ed7d494ba19b32f1d6f8b9ee4</wfw:commentRss>
            <source url="http://feeds.delicious.com/v2/rss/vreteno">vreteno's links</source>
            <description></description>
            <category domain="http://www.delicious.com/vreteno/">Federici</category>

        </item>
    </channel>
</rss>
```

ภาพที่ 24 ตัวอย่างข้อมูลที่ได้จากเว็บไซต์ Delicious.com

เมื่อได้รับข้อมูลแล้วจะทำการอ่านค่าโหนด item ทุกโหนดที่ได้รับและเลือกจัดเก็บข้อมูลลงฐานข้อมูลดังนี้

1. title คือชื่อแหล่งข้อมูลที่ผู้ใช้บริการทำการปักหมุด
2. pubDate คือวันที่ทำรายการปักหมุด
3. link คือ URL ของแหล่งข้อมูลที่ถูกปักหมุด
4. dc:creator คือชื่อของผู้ใช้บริการ
5. category คือข้อมูลแท็กของแหล่งข้อมูลที่ถูกปักหมุด (สามารถมีได้มากกว่า 1 โหนด)

ข้อมูลที่ถูกจัดเก็บจะถูกนำไปใช้ในโปรแกรมในส่วนของการจำลองการใช้งานของ

ผู้ใช้บริการในหัวข้อ 4.3.2

4.3.2 **ส่วนการจำลองการใช้งานเว็บไซต์** Delicious.com

ส่วนนี้เป็นส่วนของการจำลองการใช้งานของผู้ใช้บริการเว็บไซต์ Delicious.com เพื่อทำการสร้างแท็กคลาวด์และแท็กคลาวด์เสมือนของผู้ใช้งานบริการและแหล่งข้อมูล โดยเลือกใช้รายการการปักหมุดของผู้ใช้บริการจำนวน 147,000 รายการมาทำการจำลองการใช้งาน

ทั้งนี้ได้มีการปรับค่าขีดแบ่งเป็นค่าที่แตกต่างกันในการจำลองการใช้งานในแต่ละครั้ง เพื่อให้ได้แท็กคลาวด์และแท็กคลาวด์เสมือนที่มีจำนวนหลากหลาย โดยจะกล่าวถึงรายละเอียดของผลการทดลองในบทที่ 5

เนื่องจากมีข้อมูลนำเข้าเป็นปริมาณมาก ดังนั้น เพื่อความรวดเร็วในการจำลองการใช้งาน ผู้วิจัยได้เลือกใช้ฐานข้อมูล Mongo DB ซึ่งเป็นฐานข้อมูลแบบ No SQL แทนที่ระบบจัดการฐานข้อมูลแบบปกติโดยลักษณะโครงสร้างของข้อมูลที่เก็บจะอยู่ในรูปแบบ JSON (JavaScript Object Notation) โดยจะมีการเข้ารหัสไบนารีบนข้อมูล JSON เป็น BSON (Binary JSON)

ลักษณะของโครงสร้างฐานข้อมูลที่จัดเก็บลงใน Mongo DB แสดงดังตารางที่ 1 และตารางที่ 2

ตารางที่ 1 โครงสร้างข้อมูลที่จัดเก็บข้อมูลผู้ใช้บริการ (Author)

No.	*Field*	*Data Type*	*Description*
1	_id	Object ID	Primary Key ของผู้ใช้บริการ (ถูกสร้างโดย Mongo DB)
2	name	String	ชื่อของผู้ใช้บริการ
3	total_tag_time	long	จำนวนครั้งที่ผู้ใช้บริการเข้ามาใช้งานระบบ
4	tag_space	List<Tag>	รายการแท็กคลาวด์ของผู้ใช้บริการ
5	virtual_tag_space	List<Tag>	รายการแท็กคลาวด์เสมือนของผู้ใช้บริการ
6	tag_space_trend	String	จัดเก็บการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์
7	virtual_tag_space_trend	String	จัดเก็บการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือน

ตารางที่ 2 โครงสร้างข้อมูลที่จัดเก็บแหล่งข้อมูล (Item)

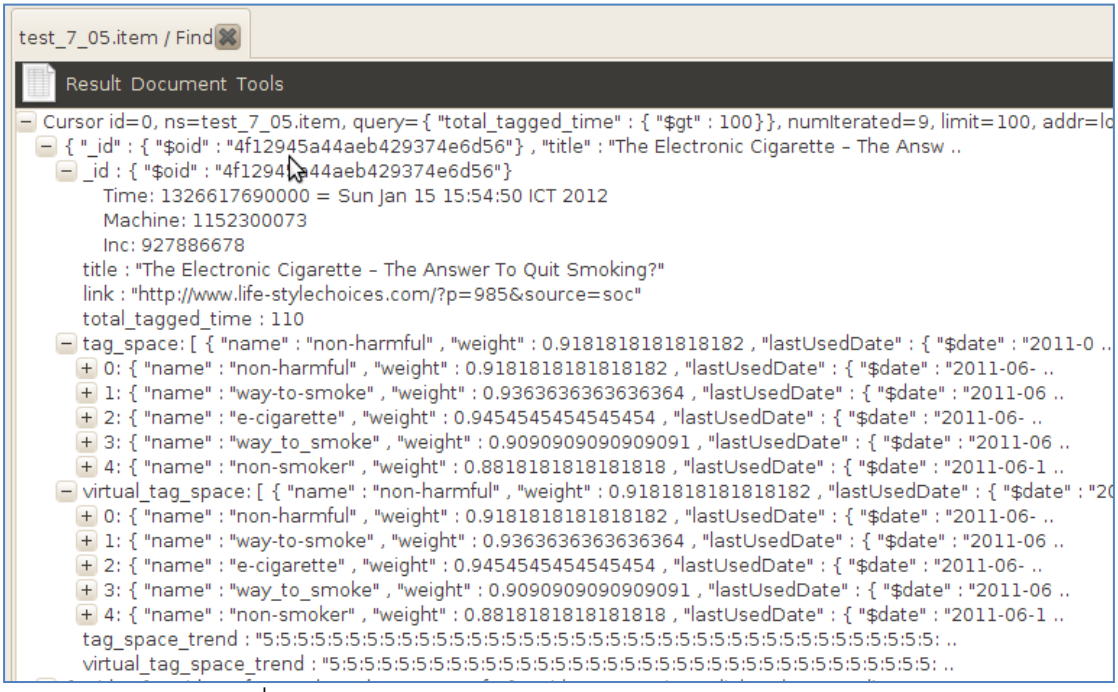
No.	*Field*	*Data Type*	*Description*
1	_id	Object ID	Primary Key ของแหล่งข้อมูล (ถูกสร้างโดย Mongo DB)
2	title	String	ชื่อของแหล่งข้อมูล
3	link	String	URL ของแหล่งข้อมูล
4	total_tagged_time	long	จำนวนครั้งที่ผู้ใช้บริการเข้ามาใช้งานระบบ
5	tag_space	List<Tag>	รายการแท็กคลาวด์ของผู้ใช้บริการ
6	virtual_tag_space	List<Tag>	รายการแท็กคลาวด์เสมือนของผู้ใช้บริการ
7	tag_space_trend	String	จัดเก็บการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์
8	virtual_tag_space_trend	String	จัดเก็บการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือน

โดยที่ลักษณะโครงสร้างข้อมูลที่จัดเก็บข้อมูลแท็กจะมีลักษณะดังตารางที่ 3 (แท็กเป็นส่วนหนึ่งของ Author และ Item)

ตารางที่ 3 โครงสร้างข้อมูลแท็กที่อยู่ใน Author และ Item

No.	*Field*	*Data Type*	*Description*
1	name	Object ID	ชื่อของแท็ก
2	weight	String	ค่าน้ำหนักของแท็ก
3	lastUsedDate	String	วันที่ถูกใช้งานครั้งสุดท้าย
4	totalUsedTime	long	จำนวนครั้งที่ถูกใช้งานทั้งหมด

หลังจากจำลองการใช้งานของผู้ใช้บริการเรียบร้อยแล้ว เมื่อลองเรียกดูข้อมูลจะได้ผลลัพธ์ที่อยู่ในรูปแบบ JSON ดังภาพที่ 25



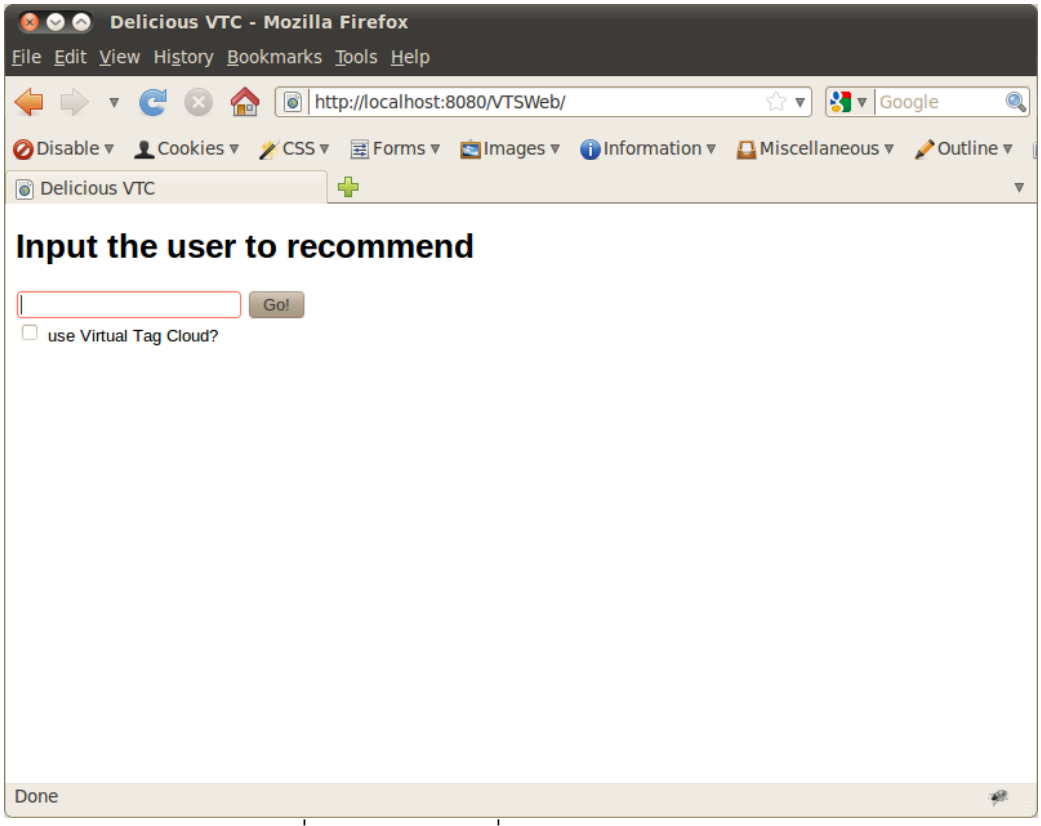
ภาพที่ 25 ตัวอย่างการเรียกดูข้อมูลแหล่งข้อมูลจากตาราง Item

4.3.3 ส่วนการแนะนำและแสดงผล

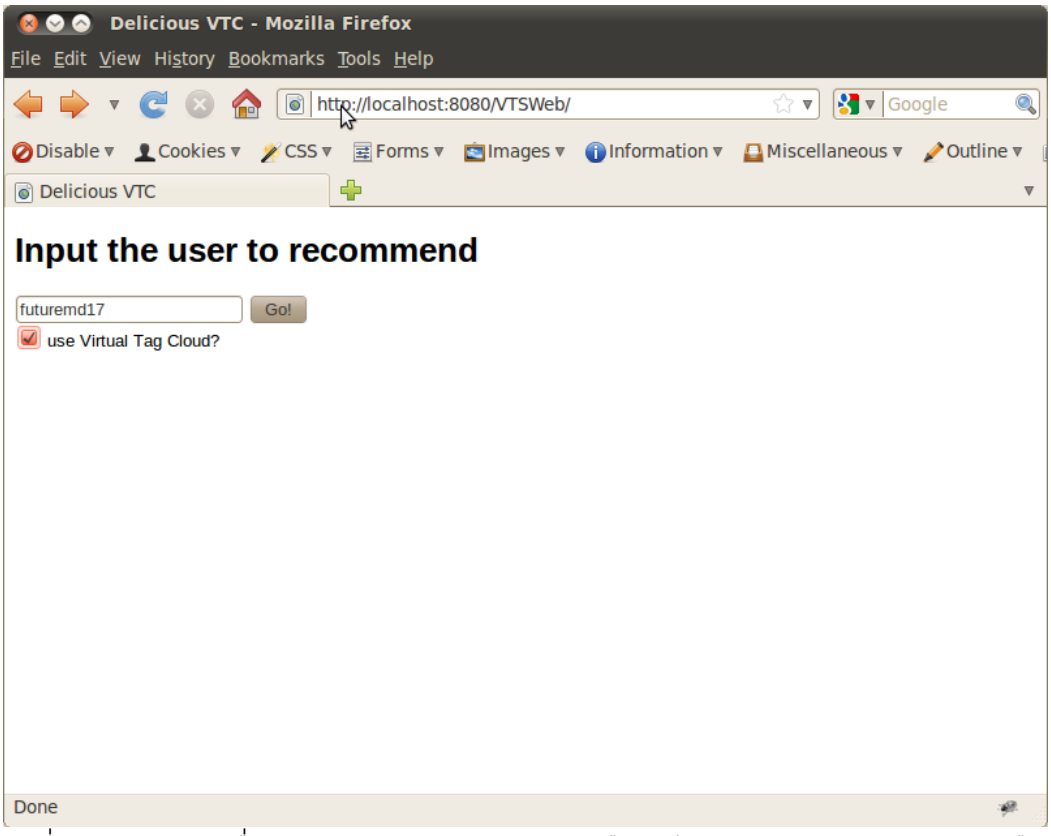
แหล่งข้อมูลที่ระบบแนะนำจะถูกนำมาแสดงบนส่วนเชื่อมต่อผู้ใช้งานในรูปแบบของเว็บแอปพลิเคชั่น โดยประกอบด้วยหน้าจอหลักๆ 3 หน้าจอคือ 1) หน้าจอเพื่อระบุชื่อเข้าใช้งาน 2) หน้าจอแสดงรายการแหล่งข้อมูลที่แนะนำ และ 3) หน้าจอแสดงการเปลี่ยนแปลงของปริมาณแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือน

4.3.3.1 หน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการ

เมื่อทำการเรียกใช้งานเว็บไซต์ ระบบจะแสดงหน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการเพื่อนำไปค้นหาแหล่งข้อมูลที่จะนำมาแนะนำดังภาพที่ 26 โดยเมื่อทำการระบุชื่อเข้าใช้งานเรียบร้อยแล้วจะสามารถเลือกได้ว่าจะใช้การแนะนำโดยใช้แท็กคลาวด์เสมือนหรือไม่ โดยสามารถคลิกที่กล่องสี่เหลี่ยมเพื่อเลือกใช้การคำนวณแบบใช้แท็กคลาวด์เสมือนดังภาพที่ 27



ภาพที่ 26 หน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการ



ภาพที่ 27 หน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการโดยเลือกวิธีคำนวณโดยใช้แท็กคลาวด์เสมือน

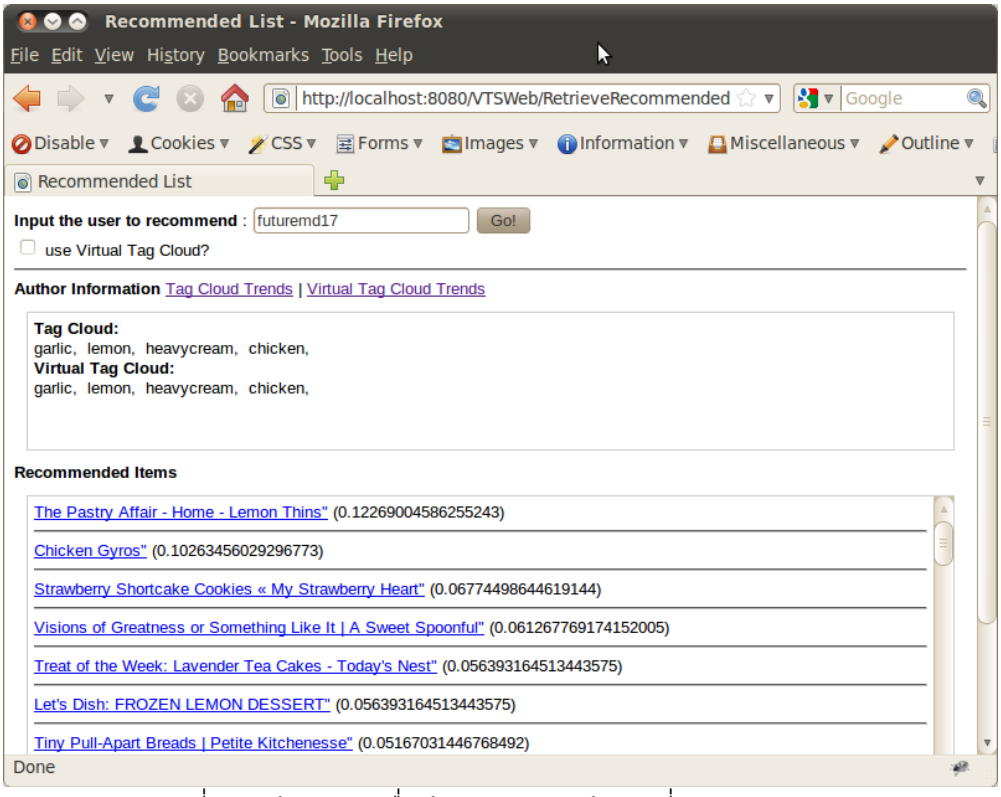
4.3.3.2 หน้าจอแสดงรายการแหล่งข้อมูลที่แนะนำ

เมื่อทำการกดปุ่ม Go จากหน้าจอระบุชื่อเข้าใช้งานของผู้ใช้บริการ ระบบจะทำการสืบค้นและคำนวณโดยวิธีการที่เลือก (แบบใช้แท็กคลาวด์หรือแท็กคลาวด์เสมือน) เมื่อทำการสืบค้นและคำนวณเสร็จแล้วจะแสดงผลดังภาพที่ 28

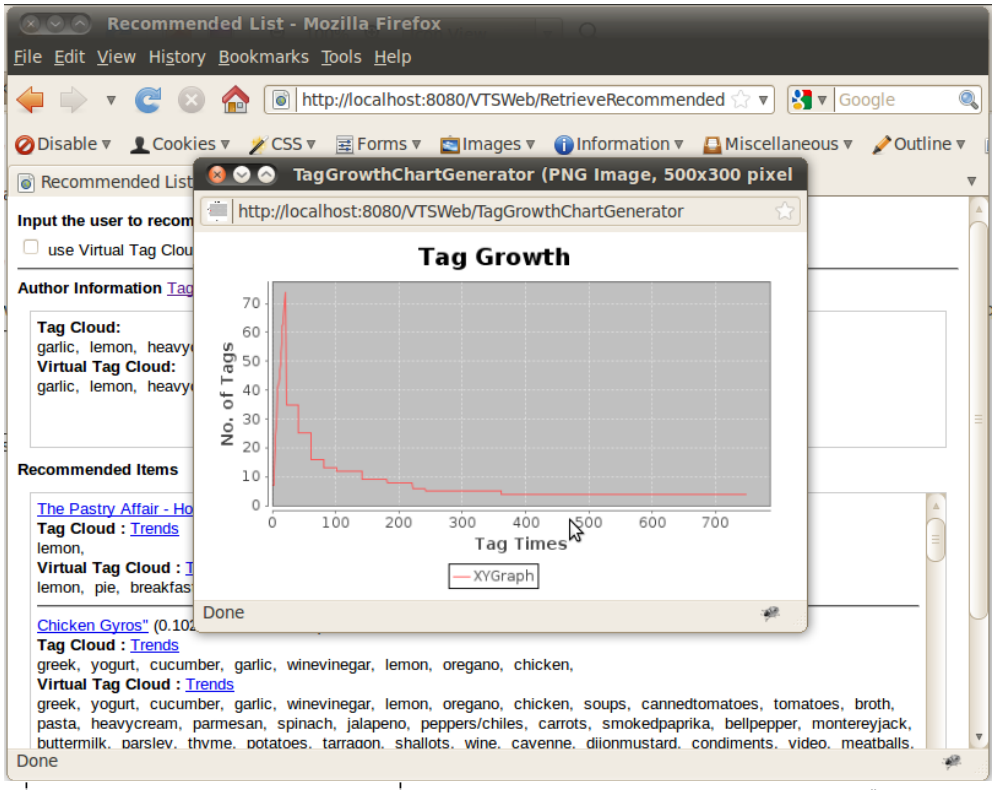
หน้าจอนี้ประกอบด้วยข้อมูล 3 ส่วนคือ

1. ส่วนการรับการป้อนชื่อเข้าใช้งานของผู้ใช้บริการ เพื่อดูรายการแหล่งข้อมูลที่ระบบแนะนำของผู้ใช้บริการอื่นๆ
2. ส่วนข้อมูลของผู้ใช้บริการ ประกอบด้วยข้อมูลแท็กภายในแท็กคลาวด์และแท็กคลาวด์เสมือน โดยจะสามารถกดลิงค์เพื่อดูแผนภูมิการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนได้ดังภาพที่ 29
3. ส่วนรายการแหล่งข้อมูลที่ระบบแนะนำซึ่งอยู่ด้านล่าง เป็นส่วนแสดงรายการแหล่งข้อมูลโดยเรียงคะแนนความคล้ายกันระหว่างผู้ใช้บริการกับแหล่งข้อมูลจากมากไปหาน้อย เมื่อทำการคลิกที่แต่ละรายการจะปรากฏข้อมูลแท็กที่อยู่ในแท็กคลาวด์และแท็กคลาวด์เสมือนดังภาพที่ 30

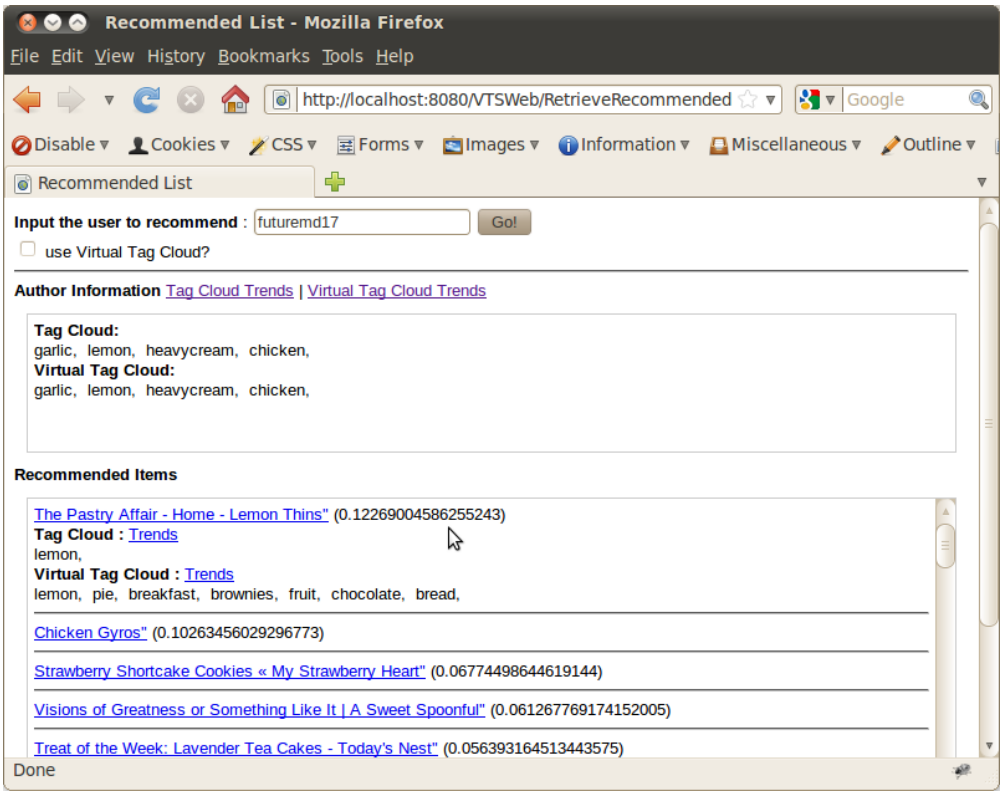
โดยจะสามารถคลิกที่ลิงค์ “Trends” เพื่อดูแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนได้เช่นเดียวกันดังภาพที่ 31



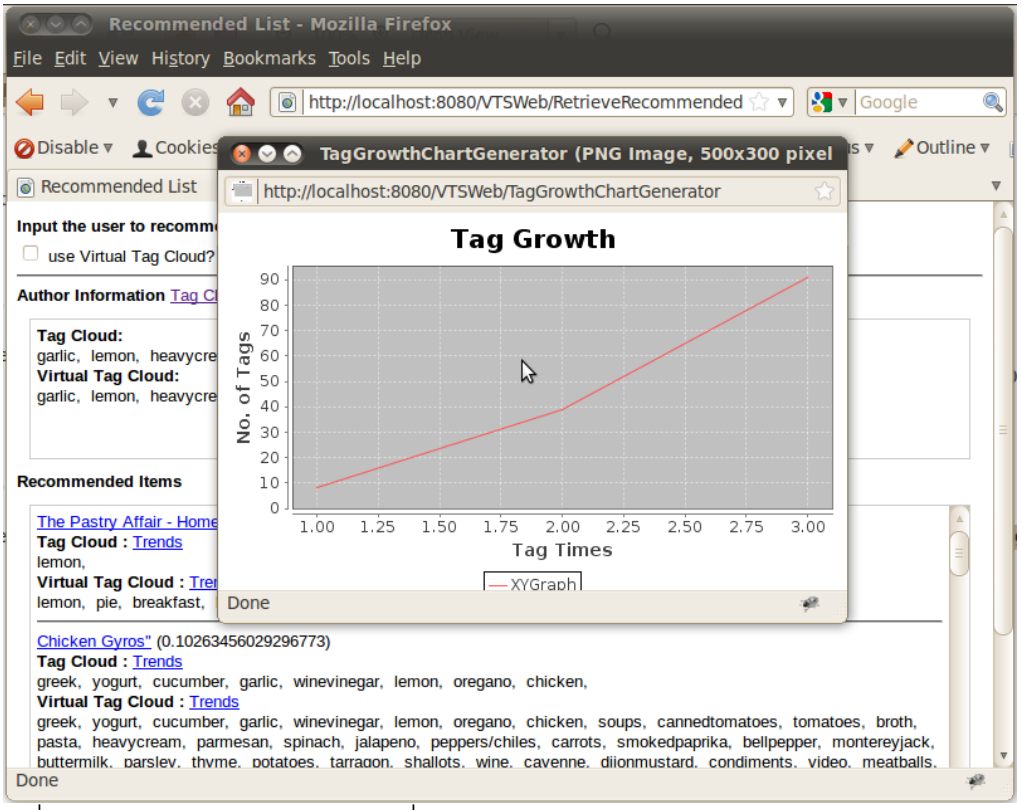
ภาพที่ 28 หน้าจอระบุชื่อเข้าแสดงแหล่งข้อมูลที่ระบบแนะนำ



ภาพที่ 29 หน้าจอแสดงแผนภูมิการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์หรือแท็กคลาวด์เสมือนของผู้ใช้บริการ



ภาพที่ 30 หน้าจอแสดงข้อมูลแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนของแหล่งข้อมูลที่ระบบแนะนำ



ภาพที่ 31 หน้าจอแสดงแผนภูมิการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์หรือแท็กคลาวด์เสมือนของแหล่งข้อมูล

บทที่ 5
การประเมินและการวัดผล

การประเมินและการวัดผลการทดลองจะใช้วิธีการคำนวณค่าความแม่นยำกับชุดข้อมูลทดสอบที่ทำการสุ่มเลือกขึ้นมาตามกลุ่มพฤติกรรมการใช้งานของประชากรในฐานข้อมูล เนื้อหาในบทนี้แบ่งออกเป็น 4 ส่วนคือ 1) วิธีการประเมินความแม่นยำของการแนะนำข้อมูล 2) การเตรียมข้อมูลเพื่อทำการทดสอบ 3) การปรับค่าพารามิเตอร์ และ 4) ผลการทดลอง

5.1 วิธีการประเมินความแม่นยำของการแนะนำแหล่งข้อมูล

การประเมินผลจะกระทำโดยการคำนวณค่าความแม่นยำของการแนะนำแหล่งข้อมูลด้วยวิธีการคำนวณค่า F-Measure หรือ F-Score ซึ่งคือค่าการวัดที่เกิดจากการคำนวณค่า Precision ร่วมกันกับค่า Recall โดยที่หากมีค่ามากจนเข้าใกล้ 1 แสดงว่าผลการแนะนำมีความแม่นยำสูง ในทางกลับกันหากค่า F-Measure มีค่าเข้าใกล้ 0 แสดงว่าผลการแนะนำมีความแม่นยำต่ำ สมการ (8) แสดงวิธีคำนวณค่า F-Measure

$$\text{F-Measure} = \mathbf{2} \cdot \frac{\mathbf{Precision \cdot Recall}}{\mathbf{Precision + Recall}} \qquad (8)$$

โดยที่ค่า Precision สามารถคำนวณได้จากสมการ (9)

$$\text{Precision} = \frac{\textbf{จำนวนแหล่งข้อมูลที่มีการแนะนำและมีความเกี่ยวข้องกับผู้ใช้บริการ}}{\textbf{จำนวนแหล่งข้อมูลที่แนะนำทั้งหมด}} \qquad (9)$$

และคำนวณค่า Recall จากสมการ (10)

$$\text{Recall} = \frac{\textbf{จำนวนแหล่งข้อมูลที่มีการแนะนำและมีความเกี่ยวข้องกับผู้ใช้บริการ}}{\textbf{จำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการทั้งหมด}} \qquad (10)$$

ยกตัวอย่างเช่น กำหนดให้

1. จำนวนแหล่งข้อมูลที่ระบบแนะนำเท่ากับ 10 แหล่งข้อมูล
2. จำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการที่มีทั้งหมดในระบบเท่ากับ 20 แหล่งข้อมูล
3. จำนวนแหล่งข้อมูลที่มีความเกี่ยวข้องกับผู้ใช้บริการเท่ากับ 1 แหล่งข้อมูล

โดยที่ จำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการคือ จำนวนแหล่งข้อมูลมีข้อมูลแท็กตรงกันกับข้อมูลแท็กของผู้ใช้บริการอย่างน้อย 1 แท็ก

จากข้อมูลข้างต้นเมื่อนำมาคำนวณค่า Precision, Recall และ F-Measure จะได้ผลลัพธ์ดังต่อไปนี้

Precision $= \frac{1}{10} = 0.1$

Recall $= \frac{1}{20} = 0.05$

F-Measure $= \frac{0.1 \times 0.05}{(0.1+0.05)} = 0.03333$

5.2 การเตรียมข้อมูลเพื่อทำการทดสอบ

จากข้อมูลรายการการใช้งานจำนวน 147,000 รายการ พบว่ามีจำนวนผู้ใช้บริการทั้งหมด 54.083 คน ภายในระยะเวลา 3 วัน จากการสังเกตผู้วิจัยได้แบ่งประชากรออกเป็น 3 กลุ่ม คือ

1. กลุ่มที่ใช้งานตามปกติ

กลุ่มของผู้ใช้บริการกลุ่มนี้คือกลุ่มผู้ใช้บริการที่เข้ามาทำรายการเป็นจำนวน 1-3 ครั้ง จากการวิเคราะห์ข้อมูลพบว่าประชากรในกลุ่มนี้มีอยู่ด้วยกัน 44,028 คน

2. กลุ่มที่ใช้งานบ่อย

กลุ่มของผู้ใช้บริการกลุ่มนี้คือกลุ่มผู้ใช้บริการที่เข้ามาทำรายการในเว็บไซต์ Delicious.com ค่อนข้างบ่อยอยู่ในช่วง 4-10 ครั้งใน 3 วัน จากการวิเคราห์ข้อมูลพบว่ามีประชากรในกลุ่มจำนวน 8,222 คน

3. กลุ่มที่ใช้งานมากกว่าปกติ

กลุ่มของผู้ใช้บริการกลุ่มนี้คือกลุ่มผู้ใช้บริการที่เข้ามาทำรายการมากกว่าปกติหรือมากกว่า 10 ครั้งภายในระยะเวลา 3 วัน จากการวิเคราะห์ข้อมูลพบว่ามีประชากรกลุ่มนี้จำนวน 1,833 คน

เมื่อทำการแบ่งกลุ่มประชากรแล้วผู้วิจัยได้สุ่มเลือกตัวแทนจากแต่ละกลุ่มประชากรมากลุ่มละ 10 เปอร์เซนต์นั่นคือ 4,402 คนจากกลุ่มแรก 822 คนจากกลุ่มที่สองและ183 คนจากกลุ่มที่สาม รวมทั้งสิ้น 5,407 คน เพื่อนำมาคำนวณค่า Precision, Recall และ F-Measure

5.3 การปรับค่าพารามิเตอร์

ในการวิเคราะห์และประเมินผลจะมีการปรับค่าพารามิเตอร์ต่างๆ ที่ใช้ในสูตรการคำนวณเพื่อหาค่าพารามิเตอร์ที่เหมาะสม ค่าพารามิเตอร์ที่สามารถปรับเปลี่ยนในงานวิจัยนี้มี 3 ค่า ดังนี้

5.3.1 ค่าขีดแบ่ง (Threshold)

ระบุถึงค่าน้ำหนักน้อยที่สุดที่แท็กในแท็กคลาวด์หรือแท็กคลาวด์เสมือนสามารถมีได้ ในงานวิจัยนี้มีการปรับค่าในช่วง 0.01-0.05 เพื่อหาค่าที่ดีที่สุด โดยจะเพิ่มขึ้นทีละ 0.01 จำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนจะแปรผกผันกับค่าขีดแบ่ง

5.3.2 **ค่าครึ่งชีวิต** (Half Life)

ระบุถึงระยะเวลาในการลืมเลือนในสมการการลืมเลือนที่อธิบายไว้ในหัวข้อ 3.2.3 สำหรับงานวิจัยนี้ได้ตั้งค่านี้เป็นค่าคงที่เท่ากับ 7 วัน เนื่องจากข้อมูลการใช้บริการของเว็บไซต์ Delicious.com ที่มีนำมาทดสอบ มีระยะเวลา 3 วัน

5.3.3 **ค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุด** (K-Nearest Neighbors)

ระบุถึงจำนวนของเพื่อนบ้านที่ระบบจะทำการสืบค้นเพื่อนำแหล่งข้อมูลที่เพื่อนบ้านเหล่านั้นเคยไปทำรายการมาเพื่อวิเคราะห์แท็กคลาวด์และแท็กคลาวด์เสมือนสำหรับใช้ในการแนะนำแหล่งข้อมูลให้กับผู้ใช้บริการเป้าหมาย ในงานวิจัยนี้มีการปรับค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุดตั้งแต่ 10-50 คน โดยจะเพิ่มขึ้นทีละ 10 คน

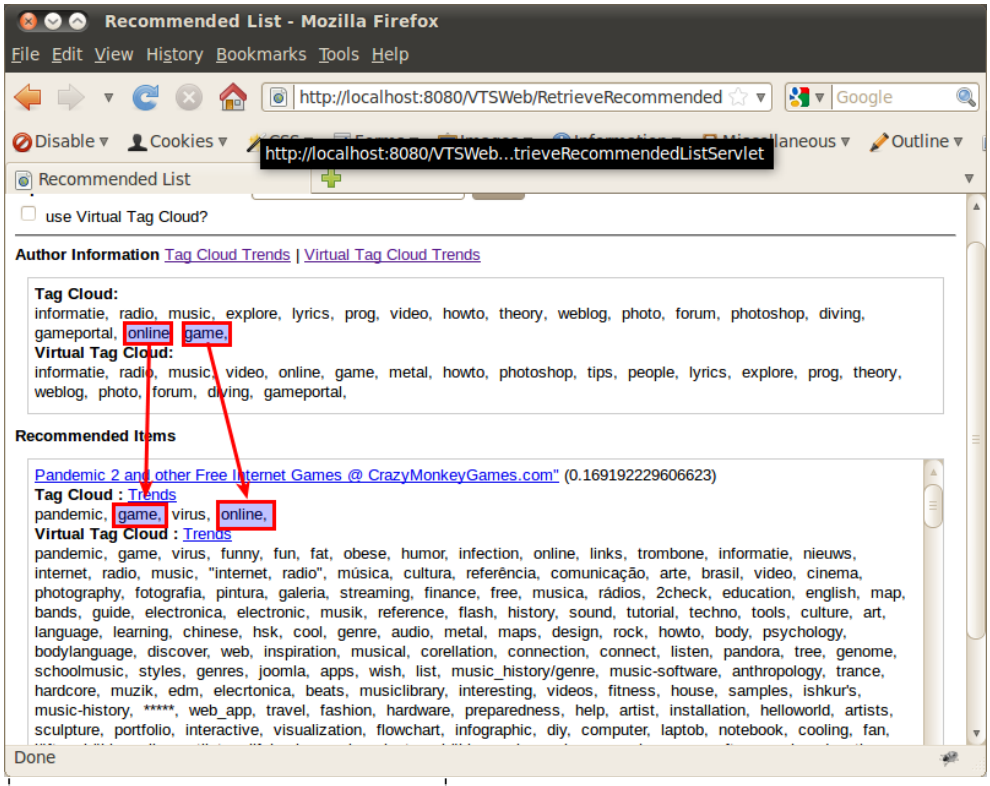
ในแต่ละชุดค่าพารามิเตอร์จะทำการคำนวณค่า Precision, Recall และ F-Measure กับตัวแทนผู้ใช้บริการจากปลุ้มประชากรที่ได้จัดกลุ่มไว้ในหัวข้อ 5.2 ที่ละคน โดยจะนำค่าเฉลี่ยที่ได้จากการคำนวณในแต่ละชุดมาสร้างแผนภูมิเพื่อวิเคราะห์และประเมินผลการทดลอง

5.4 **ผลการทดลอง**

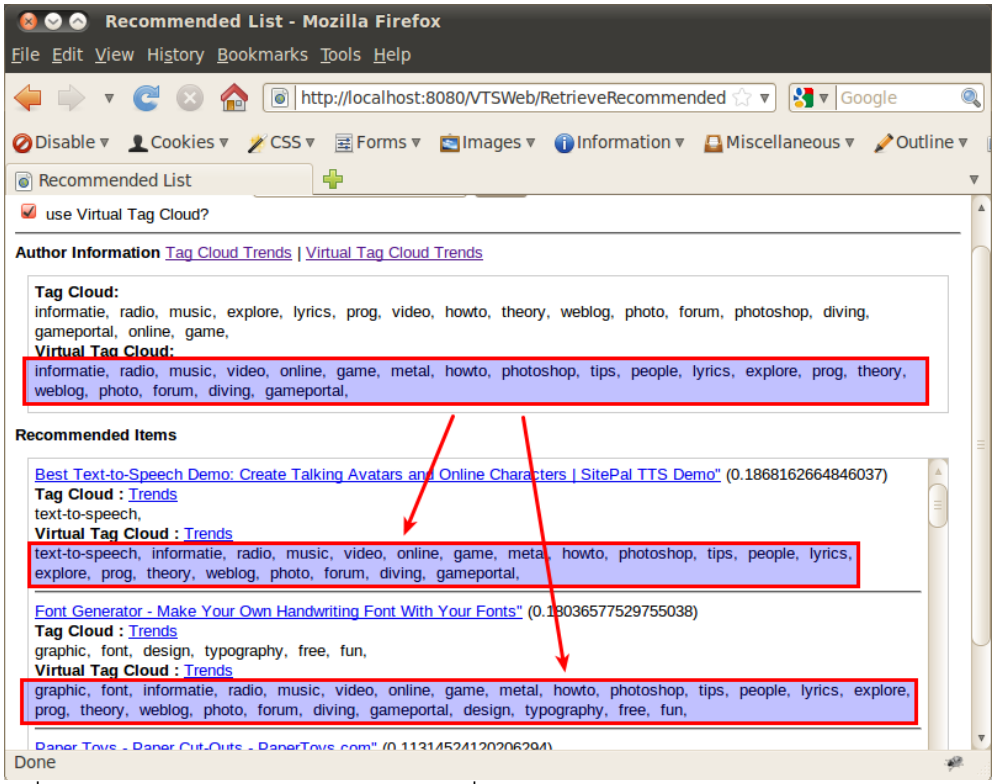
5.4.1 **รายการแหล่งข้อมูลที่ระบบแนะนำ**

ผู้วิจัยได้ทำการสังเกตรายการแหล่งข้อมูลที่ระบบแนะนำโดยวิธีการคำนวณจากแท็กคลาวด์แบบปกติและแบบแท็กคลาวด์เสมือนพบว่ารายการแหล่งข้อมูลที่ระบบแนะนำโดยใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกตินั้นจะดึงแนะนำเอาแหล่งข้อมูลที่เกี่ยวข้องกับความสนใจทางตรงของผู้ใช้บริการดังภาพที่ 32

แต่เมื่อพิจารณาแหล่งข้อมูลโดยใช้การคำนวณจากแท็กคลาวด์เสมือนจะพบว่าระบบมีการแนะนำแหล่งข้อมูลที่แตกต่างออกไปโดยจะนำข้อมูลที่มีแท็กคลาวเสมือนคล้ายกันขึ้นมาก่อนดังภาพที่ 33 ซึ่งจะสังเกตได้ว่าลักษณะของแท็กคลาวด์เสมือนของผู้ใช้บริการกับแหล่งข้อมูลนั้นจะมีความคล้ายคลึงกันอันเนื่องมาจากการแลกเปลี่ยนแท็กในระหว่างการทำรายการของผู้ใช้บริการในระบบ

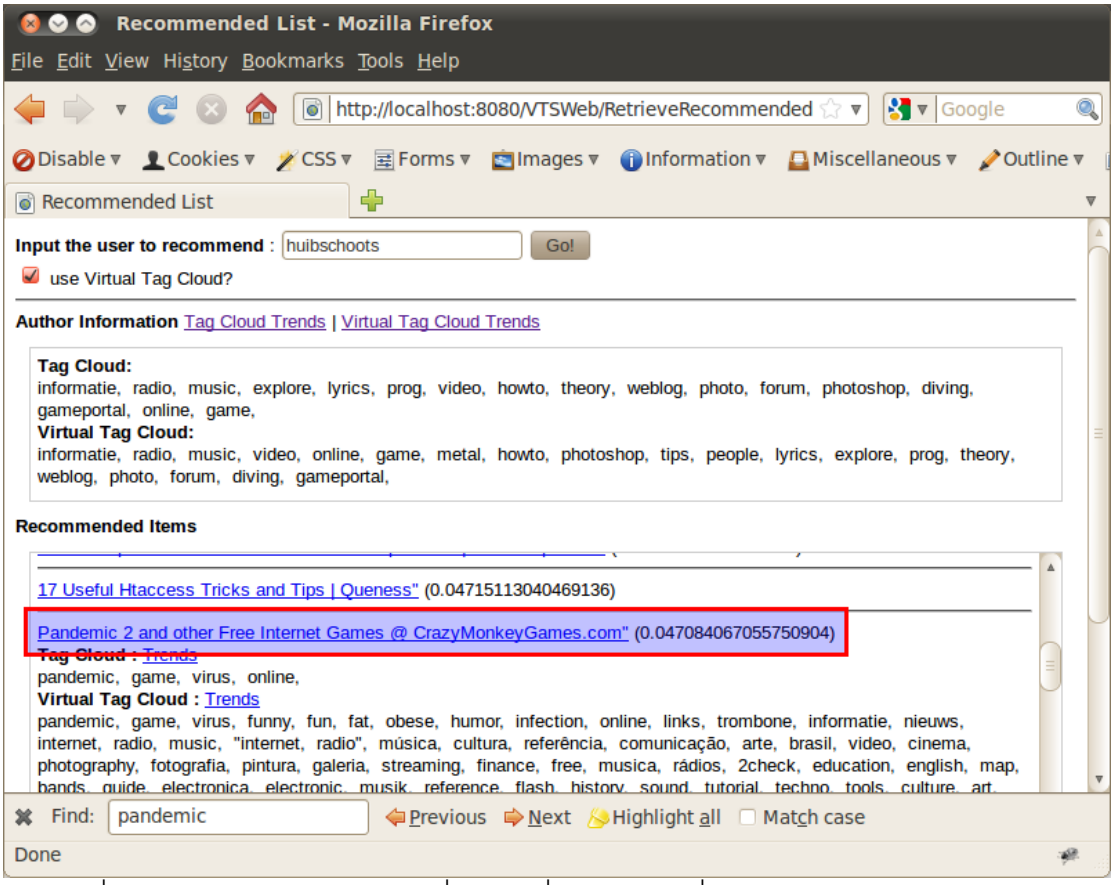


ภาพที่ 32 หน้าจอแสดงรายการแหล่งข้อมูลที่ระบบแนะนำโดยใช้การคำนวณจากแท็กคลาวด์แบบปกติ



ภาพที่ 33 หน้าจอแสดงรายการแหล่งข้อมูลที่ระบบแนะนำโดยใช้การคำนวณจากแท็กคลาวด์เสมือน

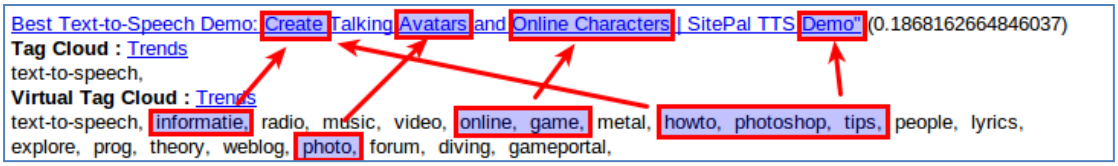
โดยที่แหล่งข้อมูลที่อยู่อันดับแรกของการแนะนำโดยใช้การคำนวณจากแท็กคลาวด์ปกติ นั้นจะปรากฏอยู่ในรายการแนะนำที่ใช้วิธีการคำนวณโดยใช้แท็กคลาวด์เสมือน แต่อยู่ในลำดับที่แตกต่างออกไปเนื่องจากค่าความคล้ายมีค่าน้อยกว่าดังภาพที่ ดังภาพที่ 34



ภาพที่ 34 ตำแหน่งของแหล่งข้อมูลที่แนะนำที่แตกต่างไปเมื่อใช้การคำนวณจากแท็กคลาวด์เสมือน

5.4.2 คุณลักษณะแฝงที่เพิ่มขึ้นในแท็กคลาวด์เสมือน

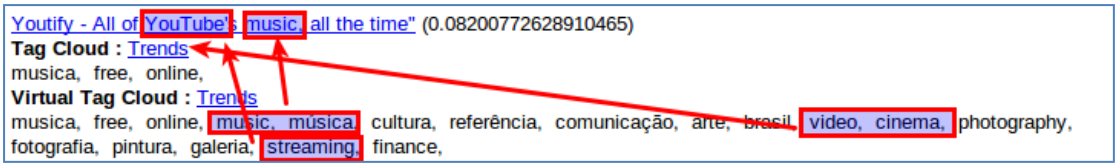
เมื่อทำการสังเกตคุณลักษณะแฝงที่เพิ่มขึ้นในแท็กคลาวด์เสมือน พบว่าคุณลักษณะแฝงที่เพิ่มขึ้นมีเกี่ยวข้องกับแหล่งข้อมูล ทำให้มีแท็กในแท็กคลาวด์เสมือนสามารถบ่งบอกถึงคุณลักษณะของแหล่งข้อมูลในมุมมองที่หลากหลายยิ่งขึ้น ดังภาพที่ 35, 36 และ และ 37



ภาพที่ 35 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล

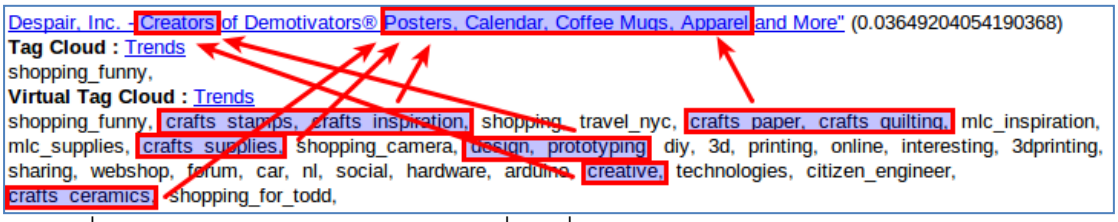
จากภาพที่ 35 จะเห็นได้ว่าแท็กที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนในขั้นตอนการแลกเปลี่ยนแท็กนั้นได้อธิบายคุณลักษณะของแหล่งข้อมูลในมุมมองที่ต่างออกไป ตัวอย่างนี้เป็น

แหล่งข้อมูลการสอนวิธีสร้างอวตาร (Avatar)ที่เอาไว้ใช้เป็นตัวแทนตัวบุคคลในการเล่นเกมหรือการใช้งานในสังคมออนไลน์ ซึ่งจะเห็นได้ว่ามีแท็ก online และ game ปรากฏอยู่ในแท็กคลาวด์เสมือนของแหล่งข้อมูล และมีแท็ก howto, photo, photoshop และ tips ซึ่งบ่งบอกว่าแหล่งข้อมูลนี้เป็นแหล่งข้อมูลที่นำเสนอวิธีการสร้างอวตารซึ่งต้องใช้รูปภาพหรือการตัดแต่งที่เกี่ยวข้องกับโปรแกรม photoshop



ภาพที่ 36 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล

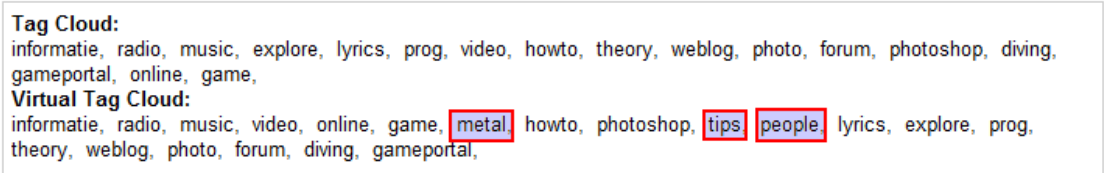
จากภาพที่ 36 จะเห็นได้ว่าแท็ก music, video, cinema และ streaming ที่อยู่ในแท็กคลาวด์เสมือนได้ช่วยบ่งบอกถึงคุณลักษณะเพิ่มเติมให้กับแหล่งข้อมูลที่เกี่ยวกับเว็บไซต์ YouTube และเนื้อหาเกี่ยวกับเรื่องดนตรี (music)



ภาพที่ 37 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของแหล่งข้อมูล

จากภาพที่ 37 จะเห็นได้ว่าแท็ก craft (งานศิลปะด้วยมือ) ประเภทต่างๆ ที่อยู่ในแท็กคลาวด์เสมือนได้ช่วยบ่งบอกถึงคุณลักษณะเพิ่มเติมให้กับแหล่งข้อมูลที่เกี่ยวกับการสร้างโปสเตอร์ ปฏิทิน ถ้วยกาแฟ และเครื่องนุ่งห่ม

เมื่อพิจารณาแท็กที่ถูกเพิ่มเข้ามาในแท็กคลาวด์ของผู้ใช้บริการพบว่าแท็กที่ถูกเพิ่มเข้าได้ช่วยทำการขยายความหรือบ่งบอกถึงความสนใจแฝงของผู้ใช้บริการเช่นเดียวกันดังภาพที่ 38



ภาพที่ 38 ตัวอย่างแท็กแสดงคุณลักษณะที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนของผู้ใช้บริการ

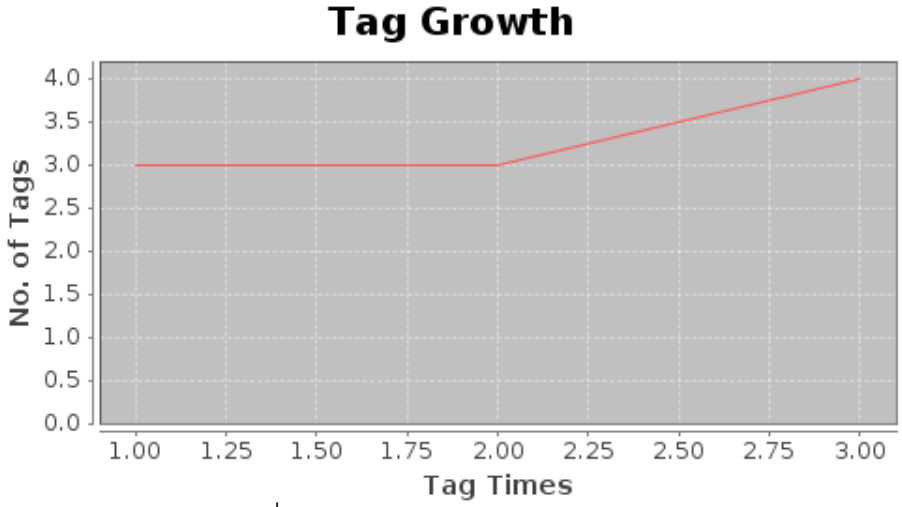
จากรูปข้างต้นจะเห็นว่ามีแท็ก metal ซึ่งช่วยบอกรายละเอียดของแนวเพลงที่สนใจ และแท็ก tips ซึ่งบ่งบอกถึงความหมายในอีกมุมมองหนึ่งของ howto และแท็ก people ซึ่งช่วยเพิ่มความหลากหลายให้กับแท็ก online และ forum

5.4.3 การเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือน

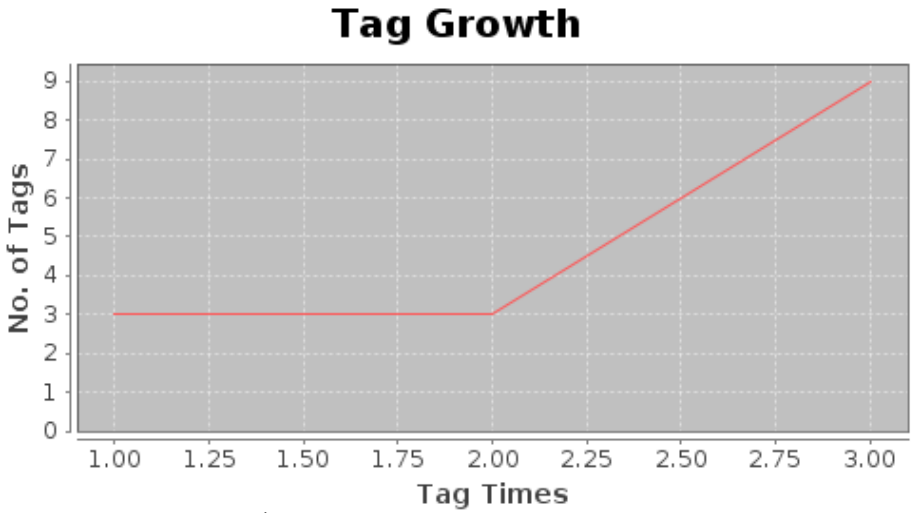
ผู้วิจัยได้ศึกษาแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนจากกลุ่มตัวอย่าง 3 ประเภทได้ผลการทดลองดังนี้

5.4.3.1 กลุ่มที่มีการใช้งานแบบปกติ

กลุ่มที่มีการใช้งานแบบปกตินั้นการเปลี่ยนแปลงจะยังไม่เห็นการเปลี่ยนแปลงที่ชัดเจน เนื่องจากจำนวนการทำรายการมีจำนวนที่น้อยเกินไป ทำให้การตัดทอนแท็กที่มีค่าน้ำหนักน้อยกว่าค่าขีดแบ่งยังไม่เกิดขึ้น ลักษณะการเปลี่ยนแปลงและจำนวนของแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนจึงมีความใกล้เคียงกันดังภาพที่ 39 และ 40



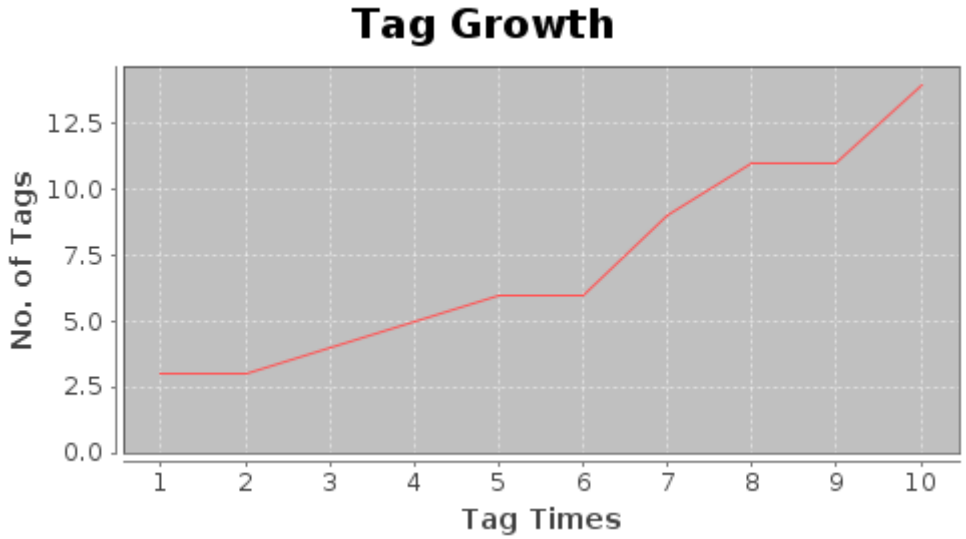
ภาพที่ 39 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานแบบปกติ



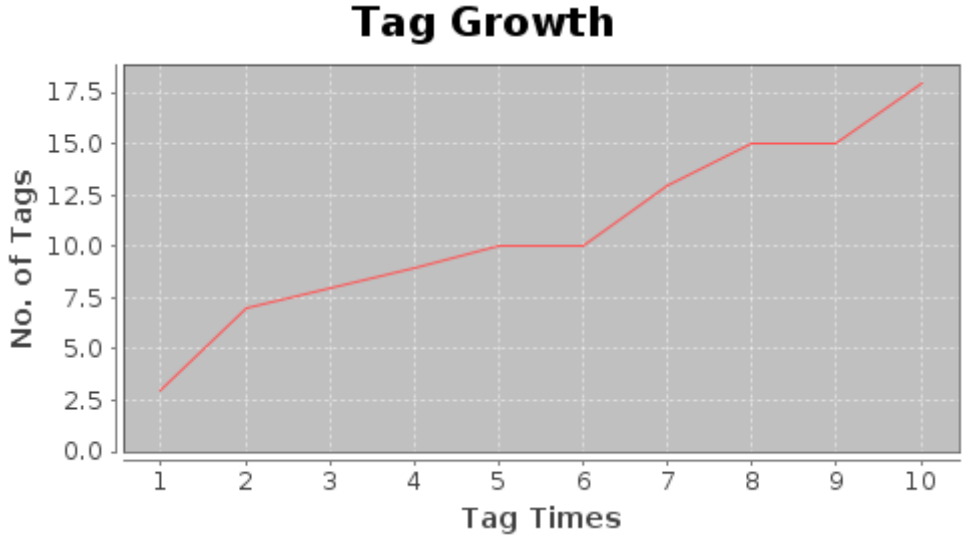
ภาพที่ 40 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานแบบปกติ

5.4.3.2 กลุ่มที่มีการใช้งานบ่อย

เมื่อทำการศึกษาการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนกับกลุ่มประชากรที่ใช้งานบ่อย (4-10 รายการ) พบว่าแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กยังเป็นไปในทิศทางเดียวกัน การตัดทอนแท็กยังไม่เข้ามาบทบาทกับจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนดังภาพที่ 41 และ 42



ภาพที่ 41 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานบ่อย

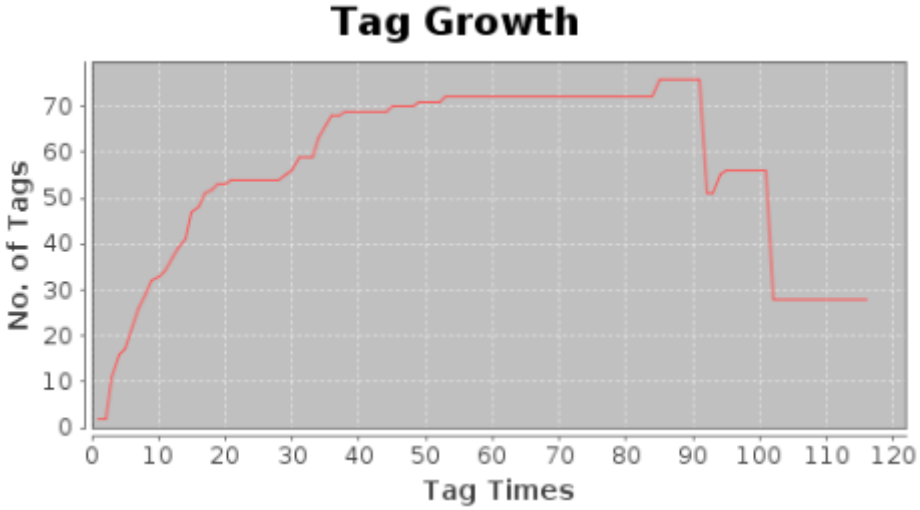


ภาพที่ 42 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนจากผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานบ่อย

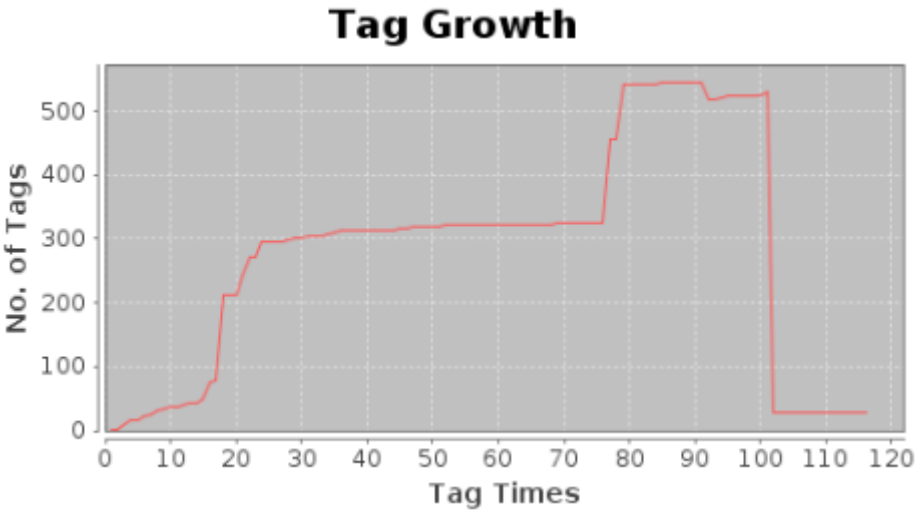
5.4.3.3 กลุ่มที่มีการใช้งานมากกว่าปกติ

เมื่อทำการศึกษาแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์

เสมือน พบว่าการแลกเปลี่ยนแท็กและการตัดทอนเข้ามาบทบาทอย่างชัดเจน โดยที่ลักษณะการเปลี่ยนแปลงจะเป็นไปในทิศทางเดียวกันคือ จำนวนแท็กจะเพิ่มขึ้นในช่วงแรกและปรับลดลงเมื่อมีการทำรายการเพิ่มมากขึ้นตามพฤติกรรมการใช้งาน แต่แตกต่างกันที่จำนวนแท็กในแท็กคลาวด์เสมือนนั้นจะมีอัตราการเพิ่มขึ้นที่รวดเร็วกว่าอันเนื่องมาจากขั้นตอนการแลกเปลี่ยนแท็ก ดังภาพที่ 43 และ 44

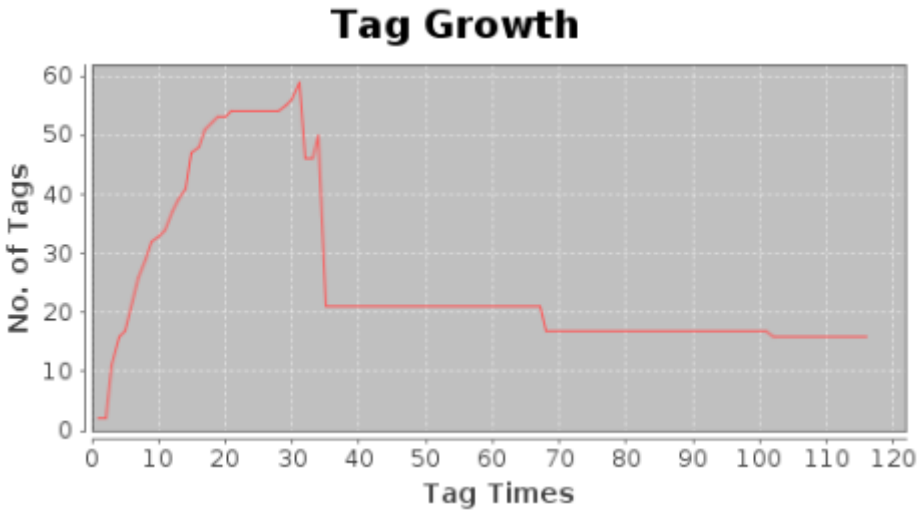


ภาพที่ 43 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานมากกว่าปกติ

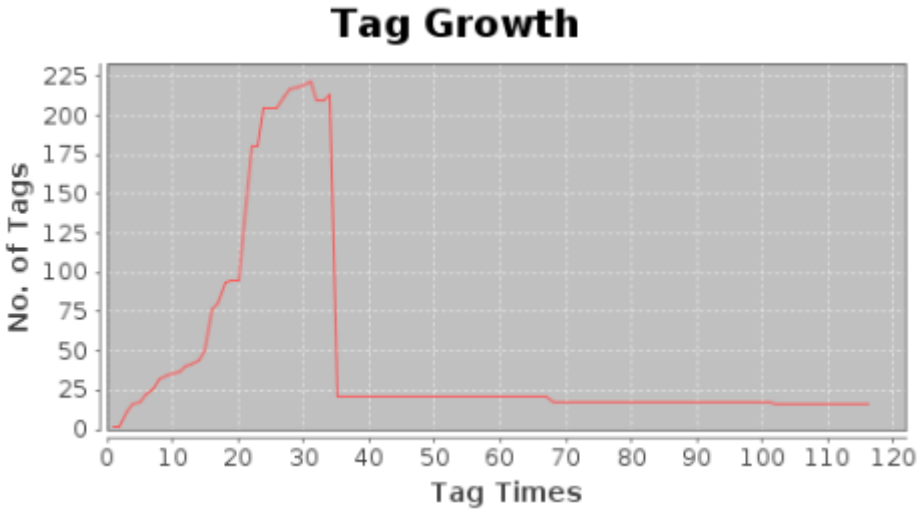


ภาพที่ 44 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการหนึ่งในกลุ่มประชากรที่ใช้งานมากกว่าปกติ

จากตัวอย่างข้างต้นแสดงแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กที่ค่าขีดแบ่งเท่ากับ 0.01 เมื่อทำการพิจารณาแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กเมื่อค่าขีดแบ่งเพิ่มมากขึ้น (0.03) พบว่าการปรับลดของจำนวนแท็กเกิดขึ้นรวดเร็วขึ้นดังภาพที่ 45 (แท็กคลาวด์) และ 46 (แท็กคลาวด์เสมือน)



ภาพที่ 45 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการเดียวกันที่ค่าขีดแบ่งเท่ากับ 0.03



ภาพที่ 46 แผนภูมิแสดงการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์เสมือนของผู้ใช้บริการเดียวกันที่ค่าขีดแบ่งเท่ากับ 0.03

เมื่อสังเกตแนวโน้มการเปลี่ยนแปลงของจำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนด้วยวิธีการคำนวณแบบต่างๆ พบว่าการแลกเปลี่ยนและการตัดทอนแท็กในแท็กคลาวด์จะส่งผลกับผู้ใช้บริการกลุ่มที่มีการใช้บริการกับระบบในปริมาณมาก โดยการปรับเปลี่ยนจำนวนแท็กจะขึ้นอยู่กับพฤติกรรมการใช้งานและค่าขีดแบ่งที่ตั้งไว้ โดยที่เมื่อค่าขีดแบ่งเพิ่มมากขึ้นจะทำให้จำนวนแท็กในแท็กคลาวด์และแท็กคลาวด์เสมือนมีการปรับลดที่รวดเร็วยิ่งขึ้น

5.4.4 ผลการคำนวณค่า Precision, Recall และ F-Measure

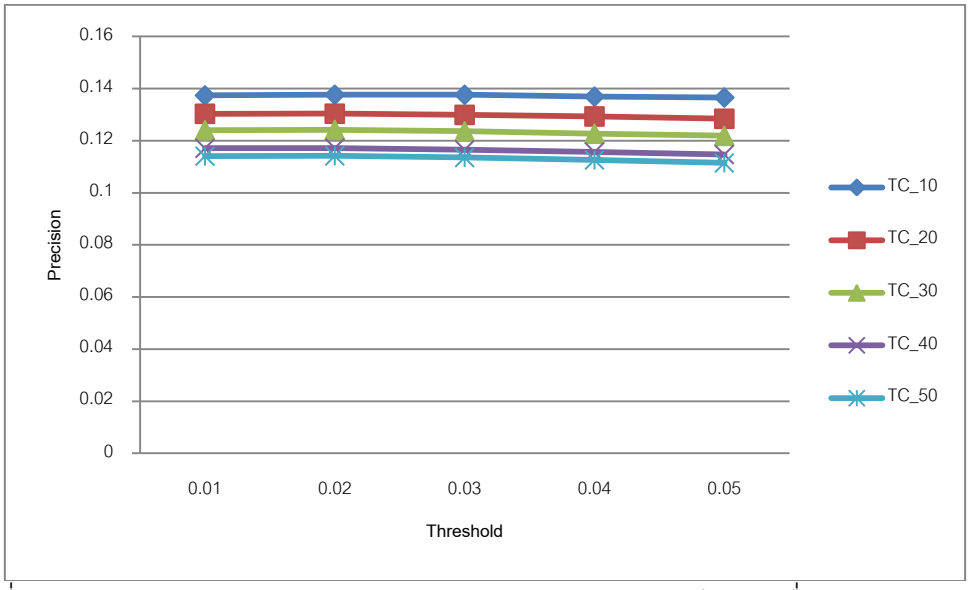
จากผลการทดลองคำนวณค่า Precision, Recall และ F-Measure จากกลุ่มตัวอย่างประชากรที่เลือกสามารถสรุปผลการวัดและการประเมินได้ดังนี้

5.4.4.1 ผลการประเมินค่า Precision

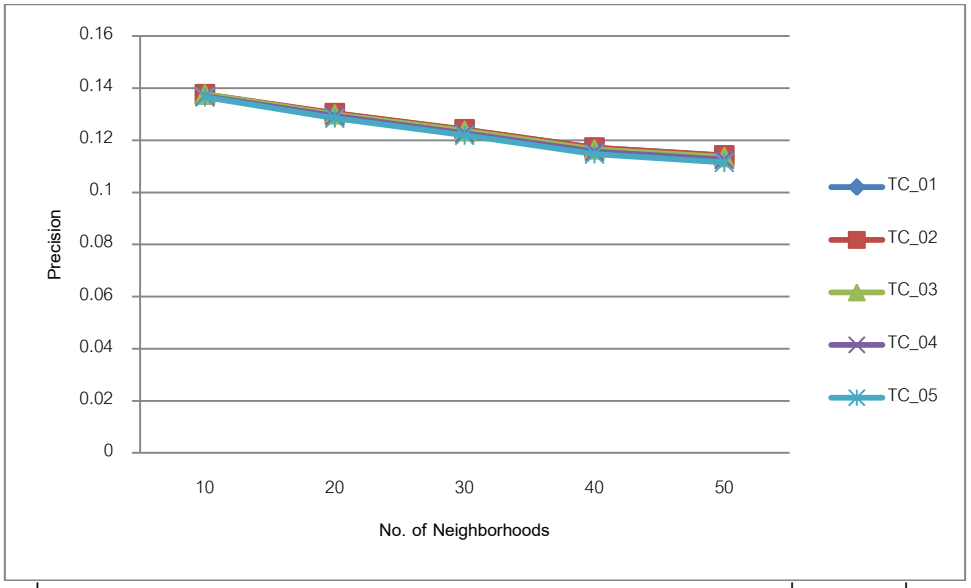
กรณีที่ใช้การคำนวณจากแท็กคลาวด์แบบปกติ การปรับค่าขีดแบ่งไม่ส่งผลทำให้ค่า Precision มีการเปลี่ยนแปลงมากนัก เนื่องจากข้อมูลแท็กในแท็กคลาวด์แบบปกติได้บ่งชี้ถึงคุณลักษณะทางตรงของผู้ใช้บริการและแหล่งข้อมูล ทำให้แหล่งข้อมูลที่ระบบแนะนำและเกี่ยวข้องกับคุณลักษณะของผู้ใช้บริการมีจำนวนใกล้เคียงกันในทุกๆ ค่าขีดแบ่งที่นำมาทดสอบดังภาพที่ 47

ในทางกลับกันเมื่อพิจารณาค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) จะพบว่าเมื่อจำนวนเพื่อนบ้านแปรผกผันกับค่า Precision ทั้งนี้เนื่องมาจากเมื่อจำนวนเพื่อนบ้านมากขึ้นจะส่งผลให้จำนวนแหล่งข้อมูลที่ระบบแนะนำมีจำนวนมากขึ้น โดยที่แหล่งข้อมูลที่แนะนำเหล่านั้นบางส่วนมีคุณลักษณะที่ไม่ตรงกับผู้ใช้บริการดังภาพที่ 48

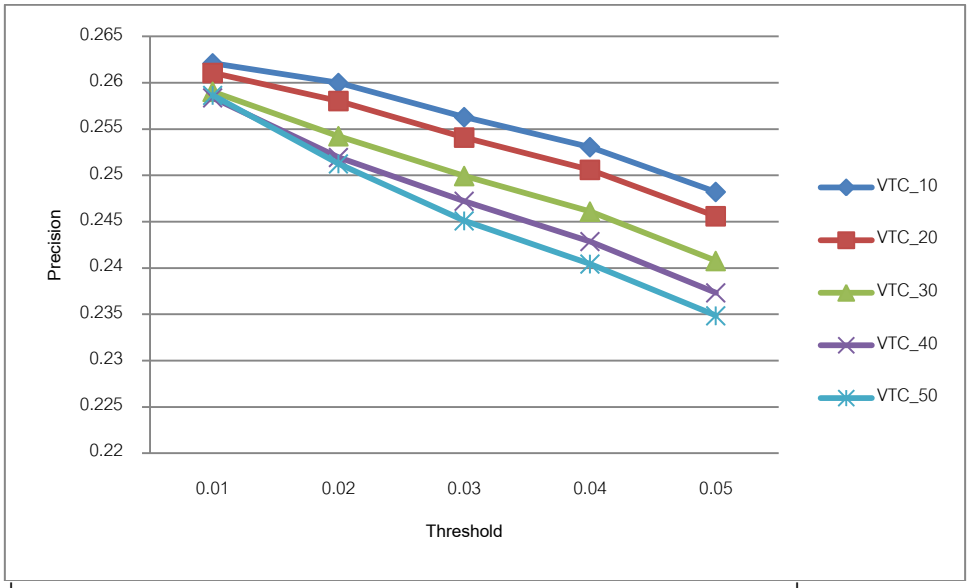
กรณีใช้การคำนวณจากแท็กคลาวด์เสมือน พบว่าค่าขีดแบ่งส่งผลกับค่า Precision โดยมีความสัมพันธ์แบบผกผันนั่นคือ หากตั้งค่าขีดแบ่งไว้น้อยค่า Precision จากมีค่ามาก ทั้งนี้เนื่องจากเมื่อค่าขีดแบ่งมีค่าน้อยจะส่งผลให้จำนวนแท็กในแท็กคลาวด์เสมือนมีจำนวนมาก ทำให้เพิ่มโอกาสที่ระบบจะแนะนำแหล่งข้อมูลที่มีความเกี่ยวข้องกับผู้ใช้บริการมากขึ้นดังภาพที่ 49



ภาพที่ 47 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ



ภาพที่ 48 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ

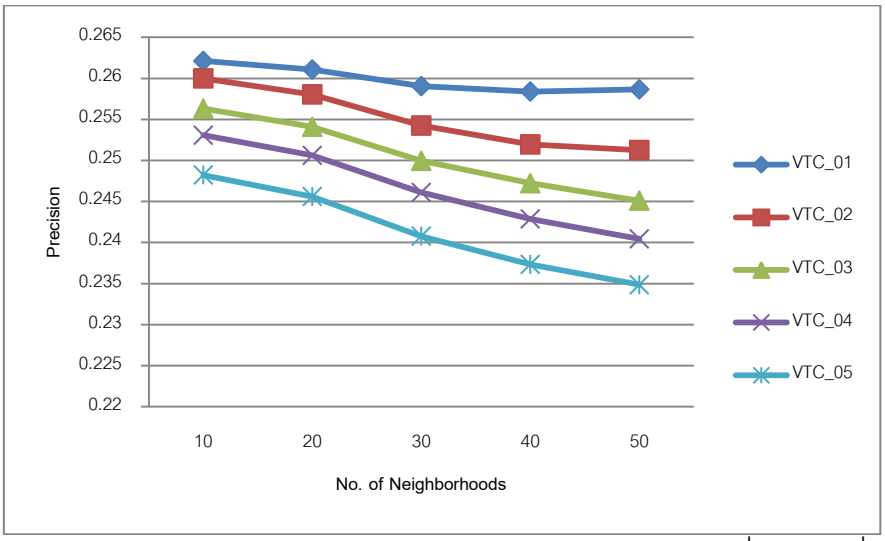


ภาพที่ 49 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน

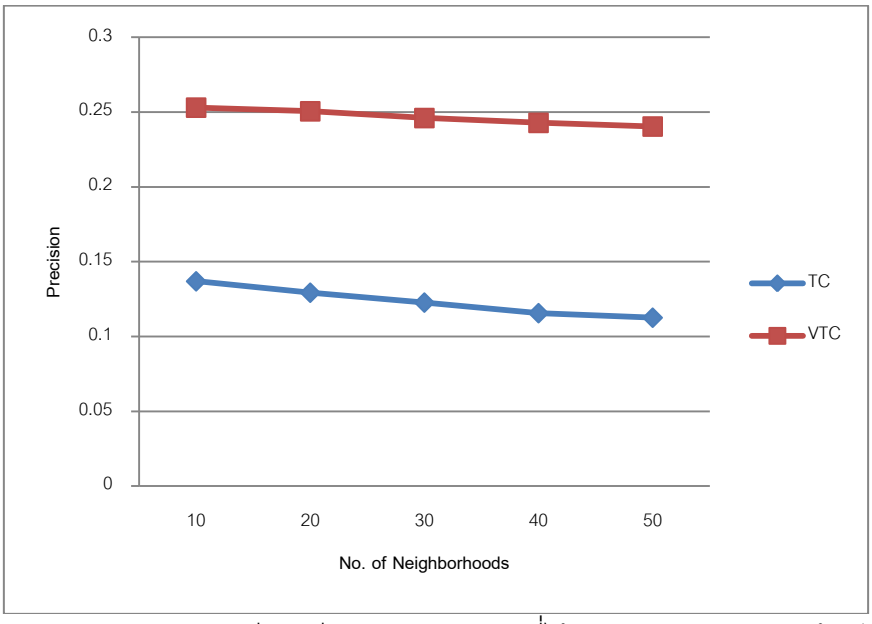
เมื่อพิจารณาค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) โดยใช้แท็กคลาวด์เสมือนในการคำนวณ พบว่าผลที่ได้มีความสัมพันธ์ระหว่างจำนวนเพื่อนบ้านกับค่า Precision เป็นความสัมพันธ์แบบแปรผกผันเช่นเดียวกับการคำนวณโดยใช้แท็กคลาวด์แบบปกติ เนื่องจากจำนวนแหล่งข้อมูลที่ระบบแนะนำและไม่เกี่ยวข้องกับผู้ใช้บริการมีจำนวนเพิ่มมากขึ้น ดังภาพที่ 50

เมื่อทำการเปรียบเทียบค่า Precision ของการคำนวณโดยใช้แท็กคลาวด์และแท็กคลาวด์เสมือน พบว่าค่า Precision ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์เสมือนมีค่าสูงกว่าค่าได้ที่จาก

การคำนวณโดยใช้แท็กคลาวด์ในทุกชุดค่าพารามิเตอร์ ดังภาพที่ 51



ภาพที่ 50 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Precision กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน



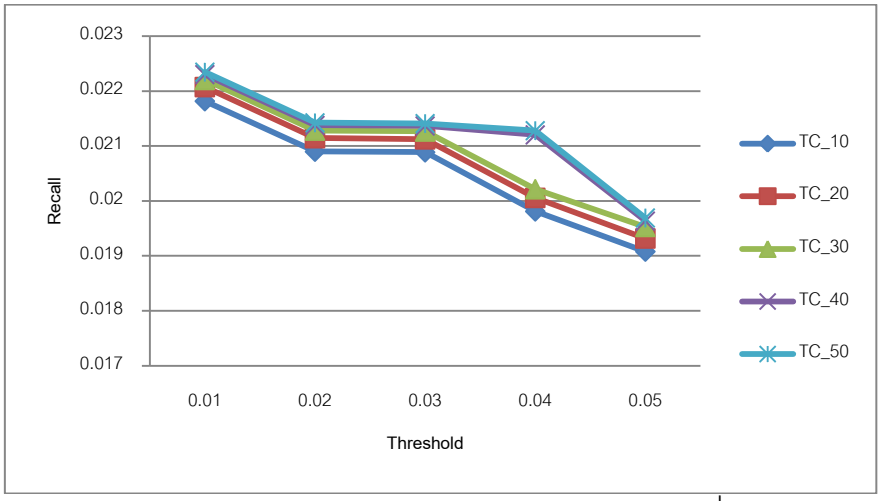
ภาพที่ 51 แผนภูมิแสดงการเปรียบเทียบค่า Precision ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน

5.4.4.2 ผลการประเมินค่า Recall

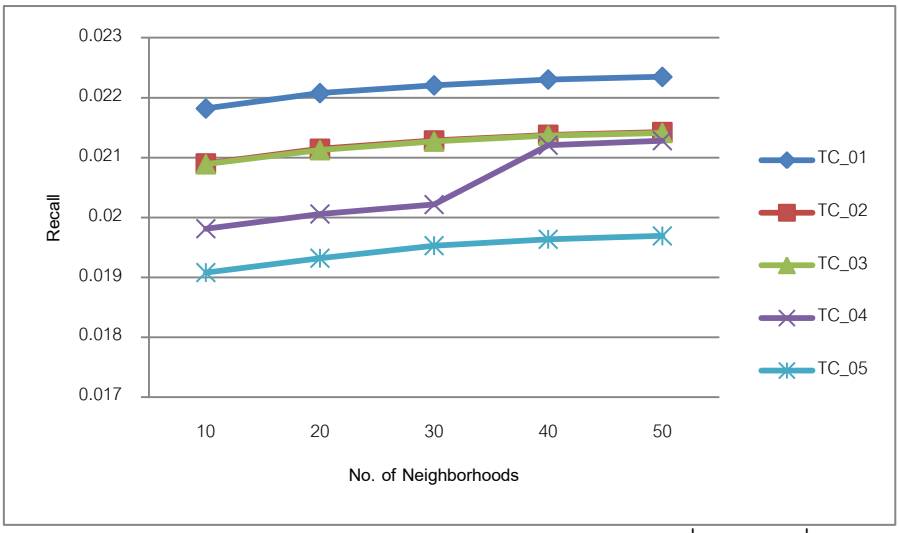
กรณีที่ใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติพบว่าค่าขีดแบ่งมีความสัมพันธ์แบบแปรผกผันกับค่า Recall เนื่องมาจากเมื่อค่าขีดแบ่งสูงจะทำให้จำนวนแท็กในแท็กคลาวด์มีจำนวนน้อยลงส่งผลให้จำนวนแหล่งข้อมูลที่ระบบแนะนำมีจำนวนน้อยเมื่อเทียบกับแหล่งข้อมูลที่เกี่ยวข้องในระบบดังภาพที่ 52

ในทางกลับกันเมื่อพิจารณาค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) จะพบว่าจำนวนเพื่อนบ้านมีความสัมพันธ์กับค่า Recall แบบแปรผันตรง เนื่องมาจากเมื่อจำนวนเพื่อนบ้านมีจำนวนมากขึ้นจำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการที่ระบบนำมาแนะนำจึงมีจำนวนมากขึ้นตามไปด้วยดังภาพที่ 53

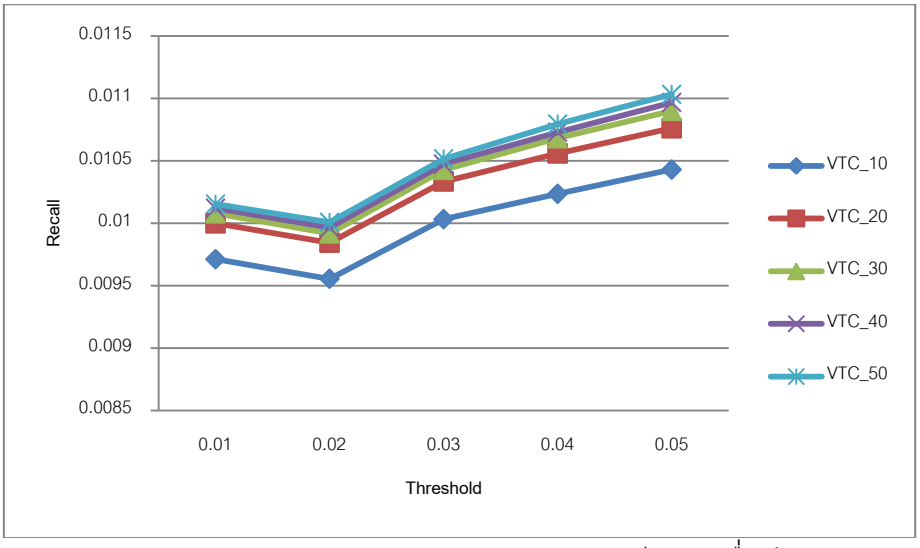
กรณีใช้แท็กคลาวด์เสมือนในการคำนวณ พบว่าค่าขีดแบ่งมีความสัมพันธ์แบบแปรผันตรงกับค่า Recall ทั้งนี้เนื่องมาจากเมื่อค่าขีดแบ่งมีค่ามากจะส่งผลให้จำนวนแท็กในแท็กคลาวด์เสมือนมีจำนวนน้อยลงทำให้จำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการมีจำนวนน้อยตามไปด้วยดังภาพที่ 54



ภาพที่ 52 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ

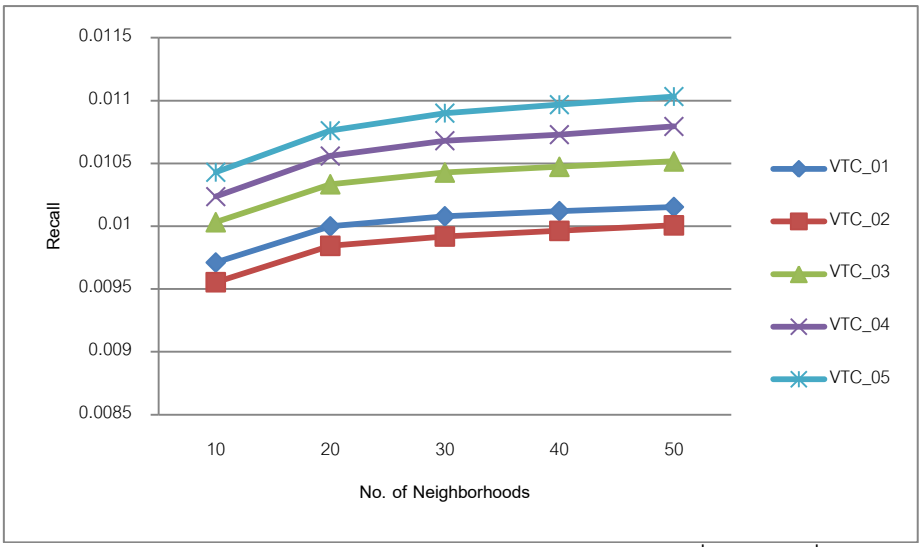


ภาพที่ 53 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ



ภาพที่ 54 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน

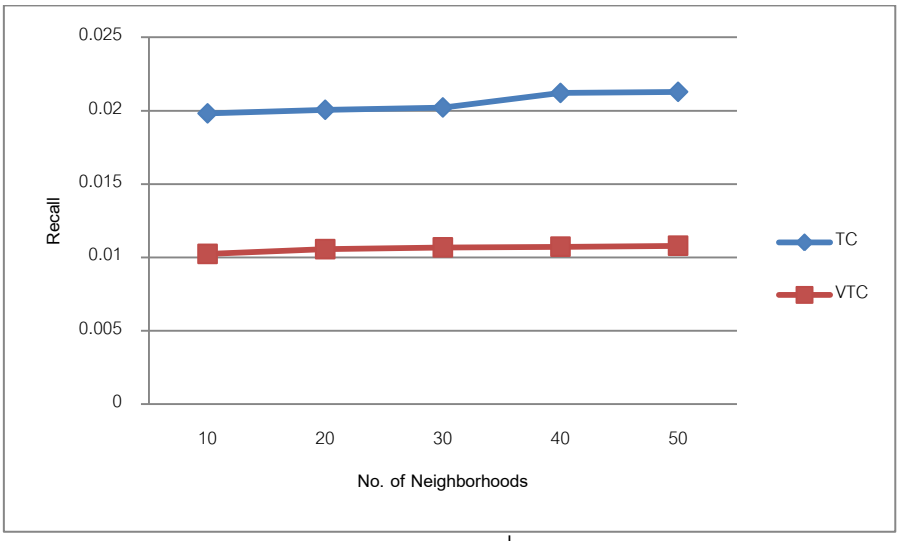
เมื่อพิจารณาค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) พบว่าจำนวนเพื่อนบ้านมีความสัมพันธ์กับค่า Recall แบบแปรผันตรงเช่นเดียวกับค่าขีดแบ่ง เนื่องจากเมื่อจำนวนเพื่อนบ้านเพิ่มมากขึ้นทำให้จำนวนแหล่งข้อมูลที่ระบบแนะนำและมีความเกี่ยวข้องกับผู้ใช้บริการมีจำนวนมากขึ้นตามไปด้วย ดังภาพที่ 55



ภาพที่ 55 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า Recall กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน

เมื่อทำการเปรียบเทียบค่า Recall ของการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือนพบว่าค่า Recall ของการคำนวณโดยใช้แท็กคลาวด์แบบปกติมีค่าสูงกว่าการคำนวณโดยใช้แท็กคลาวด์เสมือน ทั้งนี้เนื่องมาจากจำนวนแท็กในแท็กคลาวด์แบบปกติมีจำนวนน้อยกว่าจำนวนแท็กในแท็กคลาวด์เสมือน ทำให้จำนวนแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการมี

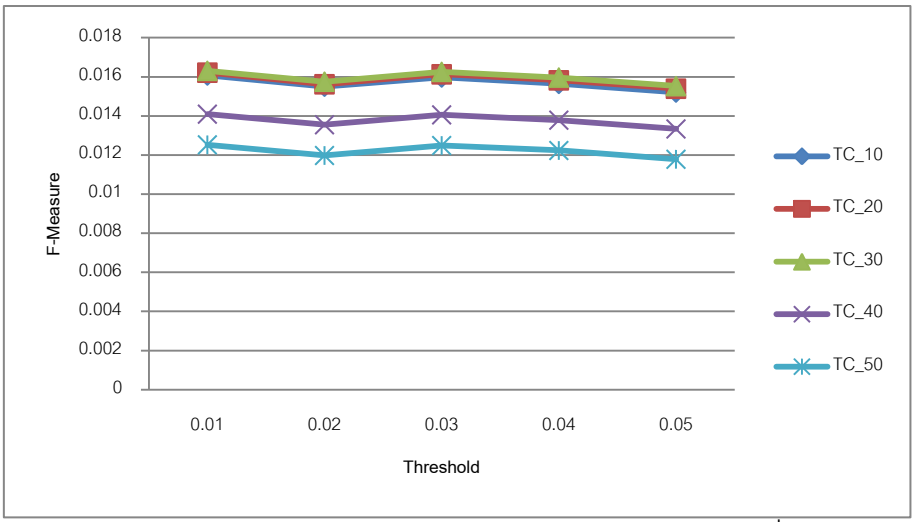
จำนวนน้อยกว่าส่งผลให้ค่า Recall มีค่าสูงกว่าดังภาพที่ 56



ภาพที่ 56 แผนภูมิแสดงการเปรียบเทียบค่า Recall ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน

5.4.4.3 ผลการประเมินค่า F-Measure

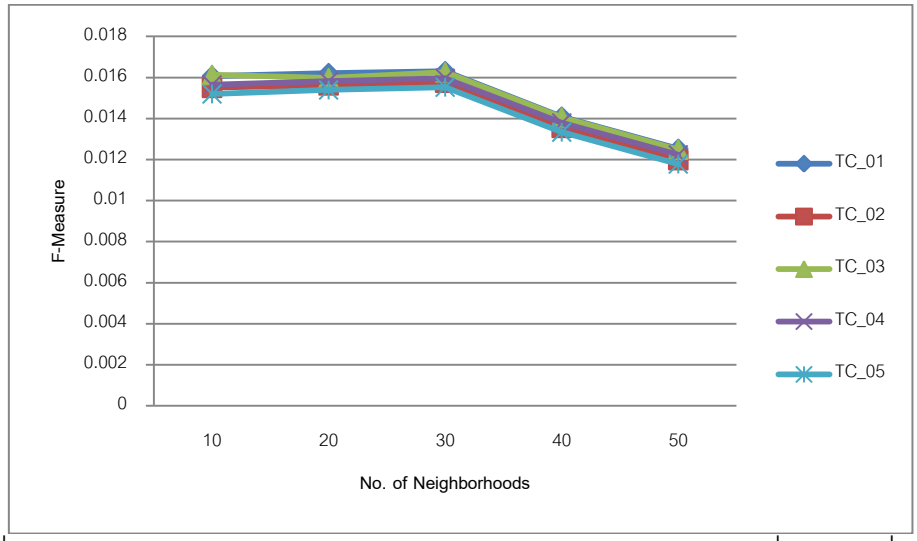
กรณีที่ใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติพบว่าค่าขีดแบ่งมีความสัมพันธ์แบบแปรผกผันกับค่า F-Measure เนื่องมาจากเมื่อค่าขีดแบ่งสูงจะทำให้จำนวนแท็กในแท็กคลาวด์มีจำนวนน้อยลงทำให้ระบบไม่สามารถนำแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการมาแนะนำได้อย่างครบถ้วนเนื่องจากจำนวนแท็กที่อยู่ในแท็กคลาวด์ของผู้ใช้บริการมีจำนวนน้อยดังภาพที่ 57



ภาพที่ 57 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ

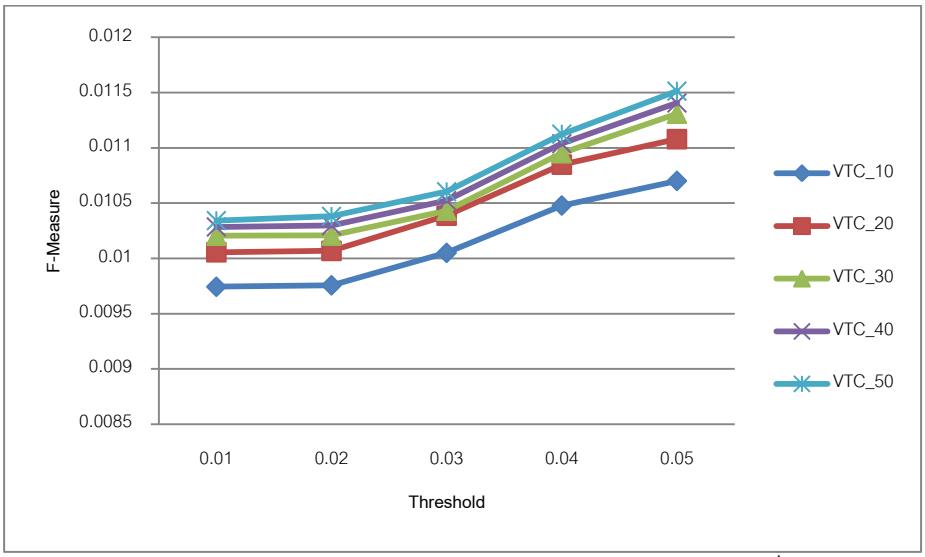
เมื่อทำการสังเกตความสัมพันธ์ระหว่างค่า F-Measure กับจำนวนเพื่อนบ้าน พบว่าจำนวนเพื่อนบ้านที่เพิ่มมากขึ้นส่งผลให้ค่า F-Measure มีค่าน้อยลงโดยเฉพาะเมื่อจำนวนเพื่อบ้าน

มีจำนวน 40-50 คนจะเห็นได้ว่าค่า F-Measure มีการปรับลดลงรวดเร็ว ทั้งนี้เนื่องจากเมื่อจำนวนเพื่อบ้านมากขึ้นส่งผลให้ระบบแนะนำแหล่งข้อมูลจำนวนมากขึ้น แต่แหล่งข้อมูลเหล่านั้นไม่มีความเกี่ยวข้องกับผู้ใช้บริการ ดังภาพที่ 58



ภาพที่ 58 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์แบบปกติ

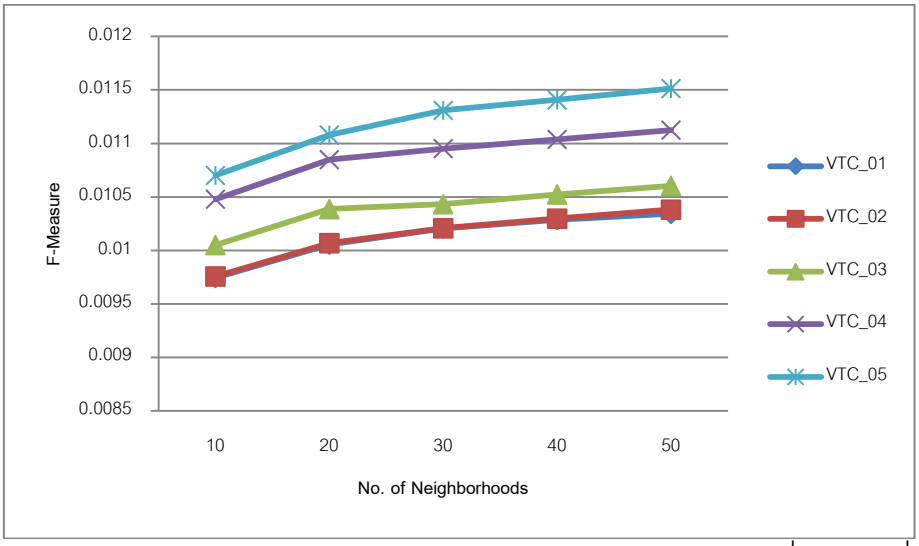
กรณีใช้แท็กคลาวด์เสมือนในการคำนวณ พบว่าค่าขีดแบ่งมีความสัมพันธ์แบบแปรผันตรงกับค่า F-Measure ทั้งนี้เนื่องมาจากเมื่อค่าขีดแบ่งมีค่ามากจะส่งผลให้จำนวนแท็กในแท็กคลาวด์เสมือนมีจำนวนน้อยลงส่งผลให้ค่า Recall มีค่าสูงขึ้นทำให้ค่า F-Measure มีค่าสูงขึ้นตามไปด้วย ดังภาพที่ 59



ภาพที่ 59 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าขีดแบ่งเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน

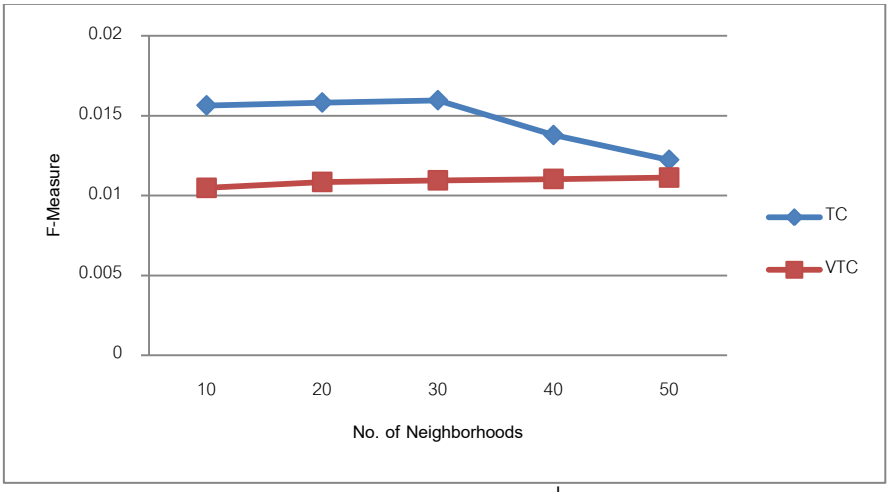
เมื่อทำการวิเคราะห์ความสัมพันธ์ระหว่างค่า F-Measure กับจำนวนเพื่อนบ้านจากการ

คำนวณโดยใช้แท็กคลาวด์เสมือน พบว่าเมื่อจำนวนเพื่อนบ้านเพิ่มขึ้นจะส่งผลให้ค่า F-Measure มีค่าสูงขึ้น ทั้งนี้เนื่องจากเมื่อจำนวนเพื่อนบ้านมากขึ้นจะส่งผลให้ระบบสามารถดึงข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการมาแนะนำได้มากขึ้น (ค่า Recall จะสูงขึ้น) ดังภาพที่ 60



ภาพที่ 60 แผนภูมิแสดงความสัมพันธ์ระหว่างค่า F-Measure กับค่าจำนวนเพื่อนบ้านเมื่อใช้การคำนวณโดยใช้แท็กคลาวด์เสมือน

เมื่อทำการเปรียบค่า F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน พบว่าค่า F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติมีค่าสูงกว่า แต่เมื่อปริมาณเพื่อนบ้านมีจำนวนสูงขึ้นจะทำให้ค่า F-Measure ลู่เข้าใกล้ค่า F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์เสมือน เนื่องจากค่า Precision ของการคำนวณโดยใช้แทกคลาวด์แบบปกติมีค่าลดลงเมื่อจำนวนเพื่อนบ้านเพิ่มมากขึ้น ในทางกลับกันค่า Recall ของการคำนวณโดยใช้แท็กคลาวด์เสมือนมีค่าเพิ่มขึ้น ดังภาพที่ 61



ภาพที่ 61 แผนภูมิแสดงการเปรียบเทียบค่า F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือน

บทที่ 6
สรุปผลการวิจัย และข้อเสนอแนะ

6.1 สรุปผลการวิจัย

เมื่อสังเกตแท็กที่ถูกเพิ่มเข้าไปในแท็กคลาวด์เสมือนจะเห็นได้ว่าแท็กเหล่านั้นได้เข้าไปเพิ่มคุณลักษณะหรือความหมายให้กับแหล่งข้อมูลและผู้ใช้บริการ โดยอาจจะอยู่ในรูปแบบของการเพิ่มมุมมองหรือความหมายแฝงให้กับแท็กที่มีอยู่แล้วในแท็กคลาวด์ให้มีความหลากหลายยิ่งขึ้น ซึ่งส่งผลกระทบให้ระบบสามารถใช้ข้อมูลแท็กเพื่อมาทำการคำนวณสำหรับการแนะนำให้มีความหลากหลายและตรงกับความต้องการทั้งทางตรงและทางอ้อมมากยิ่งขึ้น

เมื่อวิเคราะห์ค่า Precision, Recall และ F-Measure ที่ได้จากการคำนวณโดยใช้แท็กคลาวด์แบบปกติและแท็กคลาวด์เสมือนจะเห็นได้ว่า การคำนวณโดยใช้แท็กคลาวด์แบบปกติมีประสิทธิผลมากกว่าการคำนวณแบบใช้แท็กคลาวด์เสมือนเมื่อจำนวนเพื่อนบ้านมีปริมาณ 10-30 คนโดยเปรียบเทียบจากค่า F-Measure แต่เมื่อจำนวนเพื่อนบ้านเพิ่มมากขึ้นการคำนวณโดยใช้แท็กคลาวด์เสมือนจะเริ่มมีประสิทธิผลที่ใกล้เคียงกัน

ทั้งนี้เมื่อพิจารณาค่า Precision ที่ได้จากการคำนวณจะพบว่าการแนะนำแหล่งข้อมูลโดยใช้การคำนวณจากแท็กคลาวด์เสมือนให้ผลที่ดีกว่าในทุกๆ ค่าพารามิเตอร์ แต่เมื่อพิจารณาค่า Recall จะพบว่ามีค่าที่ได้จากการคำนวณโดยใช้แท็กคลาวด์เสมือนมีค่าต่ำกว่าการคำนวณโดยใช้แท็กคลาวด์แบบปกติเนื่องจากปริมาณแหล่งข้อมูลที่เกี่ยวข้องกับผู้ใช้บริการมีจำนวนมากเมื่อใช้แท็กคลาวด์เสมือน ทำให้การดึงข้อมูลขึ้นมาพิจารณาอาจจะไม่ครอบคลุมแหล่งข้อมูลที่เกี่ยวข้องทั้งหมดที่มีอยู่ในระบบ

การพิจารณาเลือกค่าขีดแบ่งและจำนวนเพื่อนบ้านที่ใกล้ที่สุดนั้น ต้องพิจารณาถึงความต้องการทางด้านจำนวนของแหล่งข้อมูลที่จะแนะนำ โดยหากต้องการให้การแนะนำมีปริมาณมากก็ควรตั้งค่าให้ค่าขีดแบ่งมีค่าน้อยๆ ซึ่งจะทำให้จำนวนแท็กที่บ่งชี้ถึงคุณลักษณะแฝงในแท็กคลาวด์และแท็กคลาวด์เสมือนถูกตัดทอนออกไปได้ยากยิ่งขึ้น หรือตั้งค่าจำนวนเพื่อนบ้านที่ใกล้ที่สุดเป็นจำนวนมาก ซึ่งส่งผลให้ระบบทำการดึงข้อมูลเพื่อนำมาพิจารณาเพิ่มมากขึ้นตามไปด้วย ทั้งนี้การปรับค่าพารามิเตอร์ดังกล่าวมีความต้องการใช้ทรัพยากรในเครื่องคอมพิวเตอร์ค่อนข้างสูงทั้งทางด้านความเร็วของการประมวลและหน่วยความจำที่ใช้งาน

จากผลการทดลองสามารถสรุปผลได้ว่าวิธีการคำนวณเพื่อแนะนำแหล่งข้อมูลโดยใช้แท็กคลาวด์เสมือนสามารถแนะนำแหล่งข้อมูลที่อยู่ในความสนใจทางตรงและทางอ้อมของผู้ใช้บริการเนื่องจากการแลกเปลี่ยนแท็กในวิธีการทำแท็กคลาวด์เสมือนได้สร้างคุณลักษณะแฝงให้กับ

ผู้ใช้บริการและแหล่งข้อมูลทำให้ปริมาณข้อมูลที่มีความเกี่ยวข้องกันในระบบมีจำนวนมากขึ้น อย่างไรก็ตาม เนื่องจากการที่ปริมาณข้อมูลที่เกี่ยวข้องมีปริมาณมากขึ้น ทำให้การสืบค้นเพื่อนำเอาแหล่งข้อมูลมาแนะนำทำได้ไม่ทั่วถึง ส่งผลให้ประสิทธิผลของการแนะนำมีค่าลดน้อยลง

6.2 ข้อจำกัด

จากการดำเนินงานวิจัยพบปัญหาและข้อจำกัดดังนี้

1. ในส่วนของการจำลองการใช้งานของผู้ใช้บริการเว็บไซต์ Delicious.com พบว่าปริมาณข้อมูลที่มีค่อนข้างมากทำให้การจำลองการทำงานเป็นไปได้ช้าและต้องใช้เนื้อที่ของหน่วยความจำและฮาร์ดดิสก์ค่อนข้างสูง ดังนั้น เครื่องคอมพิวเตอร์ที่จะนำมาทำการจำลองการใช้งานต้องมีประสิทธิภาพค่อนข้างสูง ผู้วิจัยพัฒนาส่วนจำลองการใช้งานบนระบบปฏิบัติการแบบ 64 บิท และนำฮาร์ดดิสก์แบบโซลิดสเตทเข้ามาใช้เพื่อช่วยให้กับอ่านและเขียนข้อมูลทำได้รวดเร็วยิ่งขึ้น

2. เนื่องจากปริมาณข้อมูลมีปริมาณมาก การคำนวณค่า F-Measure จากประชากรทั้งหมดทำได้ค่อนข้างยาก ผู้วิจัยได้ทำการแบ่งข้อมูลของประชากรออกเป็น 3 กลุ่มและสุ่มเลือกประชากรในแต่ละกลุ่ม กลุ่มละ 10 เปอร์เซ็นต์มาเป็นตัวแทนในการคำนวณ เพื่อลดระยะเวลาในการคำนวณ

3. วิธีการใช้แท็กคลาวด์เสมือนที่นำเสนอ เหมาะสำหรับระบบที่มีฐานข้อมูลขนาดใหญ่ที่มีการเก็บข้อมูลพฤติกรรมการใช้งานเป็นระยะเวลานาน จากฐานข้อมูลที่ผู้วิจัยได้นำมาใช้ พบว่าระยะเวลาค่อนข้างสั้น ทำให้การกระจายของประชากรไม่เป็นการแจกแจงแบบปกติ การสังเกตการเปลี่ยนแปลงของจำนวนแท็กคลาวด์จึงไม่สามารถเห็นได้ชัดเจนกับกลุ่มประชากรที่เป็นประชากรส่วนใหญ่ในฐานข้อมูล (กลุ่มใช้งานแบบปกติ) แต่จะเห็นได้ชัดเจนในกลุ่มที่ใช้งานมากกว่าปกติ ซึ่งสามารถเทียบได้กับผู้ใช้บริการที่ใช้งานมาเป็นระยะเวลานาน

4. วิธีการในงานวิจัยนี้ใช้การวิเคราะห์และคำนวณโดยอิงจากพฤติกรรมการใช้แท็กของผู้ใช้บริการ ทำให้การแนะนำแหล่งข้อมูลอาจมีความคลาดเคลื่อนในกรณีที่ผู้ใช้บริการเข้ามาใช้งานแต่ไม่มีการแท็กแหล่งข้อมูล จากการสังเกตพบว่าจำนวนแท็กในแท็กคลาวด์ของผู้ใช้บริการบางคนเป็น 0 เนื่องจากไม่เคยทำการแท็กแหล่งข้อมูลในระบบเลย

6.3 แนวทางการวิจัยต่อ

สำหรับงานวิจัยในอนาคตสามารถนำวิธีการที่นำเสนอในงานวิจัยไปใช้กับฐานข้อมูลทีมีขนาดใหญ่ขึ้นเพื่อทดสอบกับค่าครึ่งชีวิตอื่นๆ หรือประยุกต์ใช้กับฐานข้อมูลของเว็บไซต์อื่นๆ เช่น YouTube [9], CiteULike [10] หรือ facebook [11] เพื่อทำการทดลองกับลักษณะข้อมูลที่มีความ

หลากหลายยิ่งขึ้น

นอกจากนี้ ยังสามารถนำวิธีการหรือทฤษฎีการตีความหมายจากแท็กเข้ามาผนวกกับวิธีการที่นำเสนอในงานวิจัยเพื่อเพิ่มความประสิทธิผลและความแม่นยำในการแนะนำแหล่งข้อมูลให้กับผู้ใช้บริการ

รายการอ้างอิง

[1] TagClouds. Tag Cloud [Online]. 2012. Available from : http://www.tagclouds.com [2012, Jan 19]

[2] Delicious. Tag Cloud [Online]. 2012. Available from : http://www.delicious.com [2012, Jan 19]

[3] Flickr. Tag Cloud [Online]. 2012. Available from : http://www.flickr.com [2012, Jan 19]

[4] National Institute of Advance Industrial Science and Technology (AIST). Collaborative Filtering [Online]. 2012. Available from : http://www.aist.go.jp [2012, Jan 19]

[5] M. Gupta, Z. Yin, J. Han and R. Li. Survey on Social Tagging Techniques University of Techniques. SIGKDD Explorations 12 (June 2010): 58-72.

[6] N. Zheng and Q. Li. A recommender system based on tag and time information for social tagging systems. Expert Systems with Applications 38 (September 2010): 4575-4587.

[7] Y. Cheng et al. Model bloggers' interests based on forgetting mechanism. Proceeding of the 17th international conference on World Wide Web - WWW '08 (April 2008): 1129.

[8] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. Proceedings of the 18th international conference on World wide web - WWW '09 (April 2009): 641.

[9] YouTube. YouTube – Broadcast Yourself [Online]. 2012. Available from : http://www.youtube.com [2012, Jan 19]

[10] CiteULike. CiteULike:Everyone's Library [Online]. 2012. Available from : http://www.citeulike.org [2012, Jan 19]

[11] Facebook. Facebook [Online]. 2012. Available from : http://www.facebook.com [2012, Jan 19]

ภาคผนวก

ภาคผนวก ก.

ตัวอย่างตารางผลการคำนวณค่า Precision, Recall และ F-Measure

ตัวอย่างผลการคำนวณค่า Precision, Recall และ F-Measure ของกลุ่มผู้ใช้บริการบางส่วน (1 ชุดพารามิเตอร์) จากรายการคำนวณทั้งหมด 5,708 รายการ 50 ชุดพารามิเตอร์

ตารางที่ 4 ตัวอย่างผลการคำนวณค่า Precision, Recall และ F-Measure

Author	*Precision*	*Recall*	*F-Measure*
Bezobrazie	0.833	0.004	0.008
GeorgeTiffin	0.25	0.001	0.003
srnm	0.169	0.032	0.054
almondlab	0.4	0.001	0.002
jesuslatuff18	0.5	0.005	0.011
jlaskowski	0.5	0.059	0.105
regex7	0.222	0.013	0.025
doug_maclennan	1	0.006	0.011
EMoray	0.006	0.003	0.004
jana_nox	1	0.015	0.029
robynnightingale	0.557	0.041	0.076
vanghendricks	1	0.5	0.667
luigi1015	0.333	0.002	0.005
nikraz	0.3	0.001	0.001
freeurmind	0.5	0.002	0.004
stradling	0.333	0	0.001
ctv5z0t5	1	1	1
miguelmarcos	1	0.004	0.007
32flavours	0.02	1	0.038
saseda	0.2	0.002	0.004
weskus	1	0.017	0.033
Alvarpas	0.143	0	0.001
Average	0.512090909	0.123090909	0.094954545

ข้อมูลค่าเฉลี่ย Precision, Recall และ F-Measure ที่ได้จากการทดลองทั้งหมดแสดงไว้ในตารางที่ 5, 6 และ 7 โดยมีคำอธิบายถึงความหมายของตัวย่อในแต่ละรายการในแนวนอนดังนี้

ตัวอักษรภาษาอังกฤษด้านหน้า (TC หรือ VTC) แทนวิธีการที่ใช้ในการคำนวณ โดยที่

TC หมายถึงการคำนวณโดยใช้แท็กคลาวด์ และ

VTC หมายถึงการคำนวณโดยใช้แท็กคลาวด์เสมือน

ตัวเลขสองตัวด้านหลัง แทนค่าขีดแบ่งที่ใช้ในการคำนวณ โดยที่

01 เท่ากับค่าขีดแบ่ง 0.01

02 เท่ากับค่าขีดแบ่ง 0.02

03 เท่ากับค่าขีดแบ่ง 0.03

04 เท่ากับค่าขีดแบ่ง 0.04 และ

05 เท่ากับค่าขีดแบ่ง 0.05

ตัวเลข 10-50 ในแนวตั้งหมายถึงจำนวนเพื่อนบ้านที่ใกล้ที่สุดที่ใช้ในการคำนวณ

ตารางที่ 5 ตัวอย่างผลการคำนวณค่า Precision เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด

	10	*20*	*30*	*40*	*50*
TC_01	0.137424	0.130365	0.124102	0.117102	0.114104
TC_02	0.137679	0.130397	0.124205	0.117211	0.11419
TC_03	0.137647	0.129949	0.12362	0.11652	0.113553
TC_04	0.136956	0.129345	0.122739	0.115664	0.112624
TC_05	0.136603	0.128448	0.121894	0.114708	0.111466
VTC_01	0.262112	0.261069	0.25905	0.258388	0.258652
VTC_02	0.259988	0.258043	0.254247	0.251947	0.251257
VTC_03	0.256301	0.254097	0.249946	0.247224	0.245103
VTC_04	0.253064	0.250612	0.246105	0.242868	0.24044
VTC_05	0.248228	0.245603	0.240772	0.237332	0.234856

ตารางที่ 6 ตัวอย่างผลการคำนวณค่า Recall เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด

	10	*20*	*30*	*40*	*50*
TC_01	0.021821	0.022077	0.022207	0.022304	0.022348
TC_02	0.020902	0.021147	0.021285	0.02138	0.021426

TC_03	0.02089	0.021125	0.021269	0.021369	0.021412
TC_04	0.019814	0.020059	0.020216	0.021206	0.021283
TC_05	0.019081	0.01932	0.01953	0.019636	0.019693
VTC_01	0.009712	0.01	0.01008	0.010121	0.010154
VTC_02	0.009555	0.009845	0.009919	0.009964	0.010008
VTC_03	0.010033	0.010335	0.010428	0.010474	0.010517
VTC_04	0.010236	0.010559	0.010682	0.01073	0.010795
VTC_05	0.01043	0.010761	0.010901	0.010969	0.011033

ตารางที่ 7 ตัวอย่างผลการคำนวณค่า F-Measure เฉลี่ยที่ได้จากการปรับค่าพารามิเตอร์ทั้งหมด

	10	*20*	*30*	*40*	*50*
TC_01	0.016075	0.016218	0.016312	0.014095	0.012526
TC_02	0.015503	0.015624	0.015755	0.013553	0.011981
TC_03	0.015973	0.016134	0.016258	0.014059	0.012491
TC_04	0.015646	0.015825	0.015968	0.013788	0.012236
TC_05	0.015203	0.015396	0.01553	0.013338	0.011787
VTC_01	0.009744	0.010056	0.010206	0.010285	0.010342
VTC_02	0.009757	0.010068	0.010207	0.010297	0.010381
VTC_03	0.010051	0.010387	0.010433	0.010524	0.010604
VTC_04	0.010479	0.010849	0.010952	0.011038	0.011125
VTC_05	0.010702	0.011079	0.01131	0.011407	0.011514

ประวัติผู้เขียนวิทยานิพนธ์

นายคุณารักษ์ โอสถาภิรัตน์ เกิดเมื่อวันที่ 6 มกราคม พ.ศ. 2527 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2544 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552