การแนะนำเพจด้วยผู้ใช้งานผู้นำในเฟซบุ๊ก

นายเอกสิทธิ์ เพิ่มพูนพิพัฒน์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
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 University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Fanpages Recommendation Using Lead Users in Facebook

Mr. Ekasit Phermphoonphiphat

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2014

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การแนะนำเพจด้วยผู้ใช้งานผู้นำในเฟซบุ๊ก
โดย	นายเอกสิทธิ์ เพิ่มพูนพิพัฒน์
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร. บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร. บุญเสริม กิจศิริกุล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. นัทที นิภานันท์)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร. เด่นดวง ประดับสุวรรณ)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เอกสิทธิ์ เพิ่มพูนพิพัฒน์ : การแนะนำเพจด้วยผู้ใช้งานผู้นำในเฟซบุ๊ก (Fanpages Recommendation Using Lead Users in Facebook) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. สุกรี สินธุภิญโญ, 55 หน้า.

เฟซบุ๊กคือโซเชียลเน็ตเวิร์กที่ใหญ่ที่สุดของโลก มีผู้ใช้ทั่วโลกรวมกันทั้งสิ้นกว่า 1.4 พันล้านคน (สถิติเมื่อเดือนมีนาคม พ.ศ. 2558) มีผู้ใช้งานทำการเขียนเรื่องราวของตนเองลงไปทุกวัน และเฟซบุ๊กเองก็ได้มีเครื่องมือหนึ่งเรียกว่า เฟซบุ๊กเพจ ที่จะทำให้ผู้ใช้งานแบ่งปันเรื่องราวของตนเองในอีกรูปแบบหนึ่ง โดยสามารถสร้างหน้าเพจเป็นของตนเองได้ อาจจะใช้ในรูปแบบของเป็นตัวแทนของบริษัทบริษัทหนึ่ง เป็นตัวแทนของเว็บไซต์ หรือเฉพาะเรื่องราวที่ตนเองสนใจเพื่อให้ผู้ใช้งานคนอื่นติดต่อเรื่องราวของตนเอง ผู้คนโดยส่วนใหญ่จะกดไลค์เพจเพื่อติดตามเรื่องราวของเพจนั้น แม้ว่าจะกดไลค์เพจมากมายเพียงใด ผู้ใช้งานอาจไม่รู้เลยว่าอาจมีเพจที่ตรงกับความสนใจของตนเองอยู่อีก จุดประสงค์ของงานวิจัยนี้คือการที่จะแนะนำเพจให้กับผู้ใช้งาน

จุดประสงค์ของงานวิจัยนี้คือการนำเสนอวิธีการที่จะแนะนำเพจให้กับผู้ใช้งานโดยสัมพันธ์กับความสนใจของผู้ใช้งาน โดยการกดไลค์จากผู้ใช้งาน แต่เพจเองนั้นไม่มีความสัมพันธ์ระหว่างเพจด้วยกันในทางตรง ไม่เหมือนกับผู้ใช้งานทั่วไปที่มีความสัมพันธ์กันเองในลักษณะของความเป็นเพื่อน ผู้วิจัยจึงนำความสัมพันธ์ที่ผู้ใช้งานมีต่อเพจเป็นตัวเชื่อมความสัมพันธ์ระหว่างเพจ หลังจากนั้นจึงทำการแบ่งกลุ่มของเพจด้วยระยะทางจากการคำนวณการมีปฏิสัมพันธ์ร่วมกันของผู้ใช้งาน ใช้ผู้ใช้งานผู้นำในกลุ่มที่ใหญ่ที่สุดในการแนะนำเพจที่น่าสนใจ และนำเอาผลลัพธ์ที่เป็นเพจทั่วไปออกโดยนำเพจจากกลุ่มอื่นที่มีผู้ใช้งานติดตามมากที่สุดมาลบออกจากผลการแนะนำ ดังนั้นผลลัพธ์ที่เหลือ คือเพจที่แนะนำให้กับผู้ใช้งานที่ตรงกับความสนใจมากที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2557

5670472021 : MAJOR COMPUTER SCIENCE
KEYWORDS: FACEBOOK, FANPAGE, LEAD USERS, RECOMMENDATION, CLUSTERING, DISTANCE

EKASIT PHERMPHOONPHIPHAT: Fanpages Recommendation Using Lead Users in Facebook. ADVISOR: ASST. PROF. SUKREE SINTHUPINYO, Ph.D., 55 pp.

Facebook, the biggest social networks in the world, covers more than 1.41 billion monthly active users (march 2015). Many users post their story everyday. Facebook provides a way people can create their own pages call Facebook pages (previously called Fanpages). The page itself is used to represent a company, website, user's interested story, etc. Most users in facebook like some pages to follow stories they are interested, but how can they know exactly the pages they want to follow further. The goal of this research is to find a novel approach that can recommend the facebook pages for a user. Our method is based on the users' interest, the page liked by the user. The page itself does not have any relationship with other pages because it doesn’t have friends. We apply interaction with users to be relationship between pages. Then we cluster pages into the clusters and find lead users of the pages. The common pages liked by lead users are the pages we recommend to users. We avoid the popular pages that users are not interested by intersection of common pages liked by lead users from all other clusters. The pages liked by different clusters should not be an interesting ones.

Department: Computer Engineering Student's Signature .....
Field of Study: Computer Science Advisor's Signature .....
Academic Year: 2014

กิตติกรรมประกาศ

ขอขอบคุณท่านอาจารย์สุกรี สินธุภิญโญ ที่คอยประสิทธิ์ประสาทวิชาความรู้ และคอยดูแลผมในช่วงเวลาที่ผ่านมา ขอขอบคุณท่านคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ท่านอาจารย์บุญเสริม กิจศิริกุล ท่านอาจารย์นัททที นิภานันท์ และท่านอาจารย์เด่นดวง ประดับสุวรรณ ที่ให้ความกรุณาเป็นกรรมการสอบและให้คำแนะนำที่เป็นประโยชน์ต่องานชิ้นนี้ ขอขอบคุณคุณพ่อคุณแม่ พี่เอ๋ พี่อ๊อฟ เฮียอุ้ย ที่ให้การสนับสนุน ให้กำลังใจในการเรียนปริญญาโทในครั้งนี้ ขอขอบคุณต้า ที่คอยช่วยให้คำแนะนำต่างๆ เป็นเพื่อนในการปรึกษาในหลายๆเรื่อง ขอขอบคุณเพื่อนๆร่วมรุ่น CS ที่เป็นเพื่อนเรียน และปรึกษา ขอบคุณเพื่อนๆพี่ๆน้องๆที่ MIND Lab ที่ได้รับฟังการนำเสนองานวิจัยในทุกๆสัปดาห์ แลกเปลี่ยนความคิดเห็น ขอขอบคุณเพื่อนๆชมรมบัณฑิตศึกษาฯทุกคนที่ทำกิจกรรมชมรม ได้ผ่อนคลายจากการเรียนที่หนัก ขอขอบคุณอาจารย์ เพื่อนๆ พี่ๆ น้องๆ ที่ไม่ได้กล่าวถึงในที่นี้ ทุกสิ่งที่ผมประสบพบเจอในช่วงเวลา 2 ปีนี้ ส่งผลให้งานวิจัยชิ้นนี้ได้เสร็จสมบูรณ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญรูปภาพ 1

สารบัญตาราง 3

บทที่ 1 บทนำ 4

1.1 ที่มาและความสำคัญของปัญหา 4

1.2 วัตถุประสงค์ 6

1.3 ขอบเขตการดำเนินงาน 6

1.4 ขั้นตอนการดำเนินงาน 7

1.4.1 การเก็บข้อมูล 7

1.4.2 การประมวลผลข้อมูล 7

1.4.3 การสรุปผลการทดลอง 8

1.5 ประโยชน์ที่คาดว่าจะได้รับ 8

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 9

2.1 การทำคลัสเตอร์ (Clustering) 9

2.1.1 วัตถุประสงค์ของการทำคลัสเตอร์ 11

2.1.2 การแบ่งคลัสเตอร์แบบเค-มีนส์ (K-Means Clustering) 11

2.2 การวิเคราะห์โซเชียลเน็ตเวิร์ก 13

2.2.1 การวัดผลของการวิเคราะห์โซเชียลเน็ตเวิร์ก 14

2.3 งานวิจัยที่เกี่ยวข้อง 15

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

บทที่ 3 แนวคิดและวิธีการดำเนินงาน 17
3.1 หาผู้ใช้งานตัวอย่าง 17
3.2 หาเพจที่ผู้ใช้งานตัวอย่างติดตาม 21
3.3 หา 10 โพสต์ล่าสุดของแต่ละเพจ 23
3.4 หาผู้ใช้งานที่มีปฏิสัมพันธ์กับโพสต์ทั้งหมด 23
3.5 คำนวณหาระยะห่างระหว่างเพจทั้งหมด 25
3.6 การแบ่งคลัสเตอร์แบบเค-มีนส์ 29
3.7 การใช้ผู้ใช้งานผู้นำในการแนะนำเพจ 32
3.8 การแนะนำเพจโดยเปรียบเทียบกันระหว่างวิธีต่างๆ 35
3.8.1 การแนะนำโดยใช้ความชอบร่วมกัน (Common Interest Recommendation) 36
3.8.2 การแนะนำโดยใช้ผู้ใช้งานผู้นำ (Pure Lead Users Recommendation) 37
3.8.3 การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์ (Lead Users and Clusters Recommendation) 37
3.9 การวัดผล 40
บทที่ 4 ผลการทดลอง 42
4.1 ข้อมูลเพจที่ผู้ใช้งานตัวอย่างติดตาม 42
4.2 ข้อมูลโพสต์ 43
4.3 ข้อมูลความสัมพันธ์ที่เกิดขึ้น 45
4.4 ผลการทดลอง 49
บทที่ 5 บทสรุป 52
รายการอ้างอิง 53
ประวัติผู้เขียนวิทยานิพนธ์ 55

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

รูปที่ 1 เว็บไซต์ sanook.com 4
รูปที่ 2 เพจของเว็บไซต์ sanook.com 5
รูปที่ 3 ความสัมพันธ์ของบุคคลในลักษณะของเพื่อน 5
รูปที่ 4 แสดงความสัมพันธ์ระหว่างเพจโดยที่ใช้คนเป็นตัวเชื่อมความสัมพันธ์ 6
รูปที่ 5 แสดงการแบ่งกลุ่มของการแบ่งคุณลักษณะบ้าน 11
รูปที่ 6 แสดงการกระจายตัวของข้อมูล 12
รูปที่ 7 แสดงความสัมพันธ์ระหว่าโหนดโดยมีไทส์เป็นเส้นเชื่อม 13
รูปที่ 8 แสดงการเข้าไปขอโทเคน 18
รูปที่ 9 แสดงแอพพลิเคชั่นที่ผู้ใช้งานเป็นเจ้าของ 19
รูปที่ 10 แสดง App ID และ App Secret ของแอพพลิเคชั่นที่เลือก 19
รูปที่ 11 แสดงข้อมูลโพสต์ที่เรียกได้จากกราฟเอพีไอ 20
รูปที่ 12 แสดงข้อมูลที่เรียกได้จากโพสต์ 21
รูปที่ 13 แสดงข้อมูลเพจที่ผู้ใช้งานติดตามจากหน้าโปรไฟล์ 22
รูปที่ 14 แสดงโค้ดเฮชทีเอ็มแอลของเพจที่ติดตาม 22
รูปที่ 15 แสดงจำนวนผู้กดไลค์ และผู้กดคอมเมนต์ของโพสต์ 24
รูปที่ 16 แสดงหน้าจอที่สั่งให้พีเอชพีเก็บข้อมูลพร้อมกัน 25
รูปที่ 17 แสดงความสัมพันธ์ของแฟนเพจโดยใช้คนเป็นตัวเชื่อม 26
รูปที่ 18 แสดงเพจ A และ B ใกล้กัน 29
รูปที่ 19 แสดงเพจ A และ C ใกล้กัน 30
รูปที่ 20 แสดงว่าเพจ B กับ C อาจจะใกล้ หรือไกลกันก็ได้ 30
รูปที่ 21 แสดงตัวอย่างการหา centroid ใหม่ 31

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รูปที่ 22 แสดง centroid ใหม่ที่ได้จากคำนวณ 32
รูปที่ 23 แสดงผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจ 32
รูปที่ 24 แสดงการติดตามเพจร่วมกันของผู้ใช้งาน 33
รูปที่ 25 แสดงการนับจำนวนผู้ใช้งานที่ติดตามแต่ละเพจ 33
รูปที่ 26 แสดงเพจที่มีผู้ติดตามร่วมกันมากที่สุด 34
รูปที่ 27 แสดงการหาผู้ใช้งานผู้นำ 35
รูปที่ 28 แสดงเพจที่ผู้ใช้งานติดตามร่วมกัน 36
รูปที่ 29 แสดงคลัสเตอร์ที่หามาได้ 38
รูปที่ 30 แสดงการแยกคลัสเตอร์ที่หามาได้ออกเป็น 2 กลุ่ม 38
รูปที่ 31 แสดงการตั้งชื่อกลุ่มใหม่ 39
รูปที่ 32 หาเพจจากผู้ใช้งานผู้นำทั้ง 2 กลุ่ม 39
รูปที่ 33 แสดงการตัดเพจที่ไม่แนะนำออกจากผลการแนะนำ 40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ 1 ขั้นตอนการดำเนินงาน 7

ตารางที่ 2 แสดงคุณลักษณะของบ้าน 10

ตารางที่ 3 แสดงรายละเอียดของเพจที่ผู้ใช้งานตัวอย่าง 400 คนกดติดตาม 42

ตารางที่ 4 แสดงเพจที่ผู้ใช้งานตัวอย่างติดตามมากที่สุด 42

ตารางที่ 5 แสดงจำนวนปฏิสัมพันธ์ที่มีต่อเพจ เรียงจากมากไปน้อย 43

ตารางที่ 6 แสดงจำนวนปฏิสัมพันธ์ที่มีต่อโพสต์ 44

ตารางที่ 7 แสดงข้อมูลโดยรวมจากการเก็บข้อมูลทั้งหมด 45

ตารางที่ 8 แสดงตัวอย่างเพจที่ใกล้กันที่สุด 50 คู่แรก 45

ตารางที่ 9 แสดงตัวอย่างสมาชิกเพจในคลัสเตอร์ 47

ตารางที่ 10 แสดงฮิตเรตของการแนะนำเพจด้วยวิธีการทดลองทั้ง 3 รูปแบบ 49

ตารางที่ 11 แสดงฮิตเรตของการแนะนำเพจด้วยวิธีการทดลองทั้ง 3 รูปแบบ โดยเปลี่ยนสัดส่วนของเพจทดสอบ 50

ตารางที่ 12 แสดงผลลัพธ์ 10 อันดับแรก เปรียบเทียบระหว่างการใช้กับไม่ใช้คลัสเตอร์ของผู้ใช้งาน 1 คน 51

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ปัจจุบันนี้โซเชียลเน็ตเวิร์กได้เข้ามามีบทบาทในชีวิตประจำวันของคนเรามากขึ้น หลายๆคนเริ่มที่จะพูดถึงเรื่องราวของตนเอง ชีวิตประจำวัน การทำงาน เพื่อน ผ่านทางโซเชียลเน็ตเวิร์กมากยิ่งขึ้น การติดตามข่าวสารก็เช่นเดียวกัน คนเริ่มที่จะดูโทรทัศน์ อ่านหนังสือพิมพ์ หรือนิตยสารกันน้อยลง และเลือกที่จะรับข้อมูลข่าวสารผ่านทางโซเชียลเน็ตเวิร์กแทน

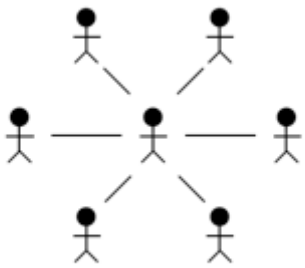
เฟซบุ๊ก (facebook) เป็นโซเชียลเน็ตเวิร์กที่มีคนใช้อยู่จำนวนมากที่สุดในโลก เปิดบริการให้ผู้ใช้งานสามารถเข้ามาอัปเดตเรื่องราวของตนเองในรูปแบบของข้อความ ภาพ และวิดิโอ อีกทั้งยังสามารถสร้างเพื่อน และติดตามเรื่องราวของเพื่อนได้ เฟซบุ๊กยังมีบริการหนึ่งที่เรียกว่า เฟซบุ๊ก เพจ (pages) โดยเปิดบริการให้ผู้ที่สนใจสามารถสร้างโพรไฟล์ใหม่เพื่อแทนตัวตนของเว็บไซต์ บุคคล ร้านขายของ บริษัท เช่นเว็บไซต์ข่าว sanook.com นอกจากจะมีเว็บไซต์เป็นของตนเองเพื่อให้บริการข่าวสารแล้ว ยังมีแฟนเพจเพื่อเป็นอีกหนึ่งช่องทางในการให้บริการข่าวสาร ทำให้ผู้ใช้งานมีช่องทางในการติดตามข่าวสารที่มากขึ้นดังตัวอย่างในรูปที่ 1 และรูปที่ 2

รูปที่ 1 เว็บไซต์ sanook.com

รูปที่ 2 เพจของเว็บไซต์ sanook.com

เราสามารถติดตามข่าวสารของเพจได้โดยการกดไลค์ (like) เมื่อแฟนเพจอัปเดตข่าว ข่าวสารนั้นๆจึงปรากฏให้เห็นได้จาก ไทม์ไลน์ (timeline) ของตนเอง หรือแม้แต่เพื่อนของเราได้ไปกดไลค์หรือ คอมเมนต์ (comment) เราก็จะมองเห็นด้วยเช่นกัน ซึ่งเป็นวิธีการที่สะดวกมากในการรับข้อมูลข่าวสารของแฟนเพจต่างๆ แต่เนื่องจากมีเพจจำนวนมากในปัจจุบัน จึงเป็นการยากที่จะเข้าถึงเพจที่น่าสนใจ และตรงกับความต้องการ จึงต้องหาวิธีในการค้นหาความสัมพันธ์ เพื่อหาข่าวสารที่ตรงกับความสนใจมากที่สุด

เพื่อน คือรูปแบบหนึ่งของความสัมพันธ์ระหว่างบุคคล โดยที่เฟซบุ๊กได้เปิดให้มีการเชื่อมความสัมพันธ์แบบเพื่อนได้ ส่งผลให้สามารถติดตามข้อความต่างๆของเพื่อนได้จากไทม์ไลน์ได้ง่ายขึ้น อีกทั้งยังมีกิจกรรมต่างๆที่สามารถทำได้เมื่อเป็นเพื่อนกันดังแสดงในรูปที่ 3



รูปที่ 3 ความสัมพันธ์ของบุคคลในลักษณะของเพื่อน

เช่นเดียวกันกับความสัมพันธ์ระหว่างเพจและบุคคล ซึ่งคนๆหนึ่งสามารถติดตามเพจที่สนใจได้เป็นจำนวนมาก เพื่อติดตามข่าวสารที่เกิดขึ้นของแต่ละแฟนเพจ ดังนั้นเมื่อพิจารณาถึงความสัมพันธ์ระหว่างเพื่อน และเพจ จึงเกิดทำให้เกิดความสัมพันธ์ระหว่างเพจขึ้นโดยใช้คนเป็นตัวเชื่อมความสัมพันธ์ดังรูปแสดงในรูปที่ 4

รูปที่ 4 แสดงความสัมพันธ์ระหว่างเพจโดยที่ใช้คนเป็นตัวเชื่อมความสัมพันธ์

ดังนั้นการบ่งบอกว่าเพจใดสัมพันธ์กับเพจใด หรือมีความใกล้เคียงกันเพียงใด เราสามารถใช้ความสัมพันธ์รอบข้างโดยใช้ผู้ใช้เป็นตัวเชื่อมความสัมพันธ์ โดยการศึกษาในครั้งนี้ได้อาศัยความสัมพันธ์ในรูปแบบดังกล่าวเพื่อหาความสัมพันธ์ในรูปแบบต่างๆระหว่างเพจ ซึ่งจะนำเสนอในหัวข้อแนวคิดและวิธีดำเนินงานอย่างละเอียดในลำดับถัดไป

1.2 วัตถุประสงค์

ศึกษาวิธีการแนะนำเพจที่ตรงกับความสนใจของผู้ใช้งานเฟซบุ๊ก

1.3 ขอบเขตการดำเนินงาน

1.3.1 การเก็บข้อมูลเพจของผู้ใช้งานแต่ละคนจะเก็บได้เฉพาะเพจที่ตั้งค่าให้มองเห็นได้เป็นสาธารณะของผู้ใช้งานเท่านั้น

1.3.2 ด้วยข้อจำกัดของการเก็บข้อมูลจาก กราฟเอพีไอ (Graph API) เมื่อเรียกข้อมูล ผู้กดไลค์ หรือ ผู้กดคอมเมนต์ มากเกินกว่า 1 ล้านคน จะเกิดข้อจำกัดที่บางครั้งไม่อาจเรียกข้อมูลต่อไปได้ จึงต้องหยุดการเรียกข้อมูลในส่วนนี้ไปทั้งที่จริงๆมีข้อมูลมากกว่านี้

1.4 ขั้นตอนการดำเนินงาน

การดำเนินงานวิจัยนี้เพื่อทดสอบการแนะนำเพจให้กับผู้ใช้งานเฟซบุ๊ก แบ่งออกเป็น 3 ขั้นตอนหลักๆด้วยกันคือ การเก็บข้อมูล การประมวลผลข้อมูล และการวัดผล ซึ่งในบทนี้จะกล่าวนำพอสังเขป การทำงานในแต่ละขั้นตอนจนได้ผลการทดลองอย่างละเอียดจะกล่าวถึงในบทที่ 3 ขั้นตอนดำเนินงานแสดงในตารางที่ 1

ตารางที่ 1 ขั้นตอนการดำเนินงาน

ขั้นตอนการดำเนินงาน	เดือน						
	1	2	3	4	5	6	7
เขียนโปรแกรมเพื่อเก็บข้อมูล	■						
เก็บข้อมูล		■	■				
เขียนโปรแกรมเพื่อประมวลผลข้อมูล			■	■			
ประมวลผลข้อมูล					■	■	
สรุปผลการทดลอง							■

1.4.1 การเก็บข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้ ทั้งหมดมาจากการเรียกข้อมูลของเฟซบุ๊ก ในขั้นตอนนี้มีจุดประสงค์เพื่อเชื่อมความสัมพันธ์ระหว่างเพจเข้าด้วยกันโดยมีผู้ใช้งานเป็นตัวเชื่อมความสัมพันธ์ ซึ่งมีการเก็บข้อมูลต่างๆดังต่อไปนี้

1.4.1.1 ข้อมูลเพจของผู้ใช้งาน

1.4.1.2 ข้อมูลโพสต์ของแต่ละเพจ

1.4.1.3 ข้อมูลผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจ

1.4.2 การประมวลผลข้อมูล

ขั้นตอนนี้ใช้การทำคลัสเตอร์ (cluster) มาช่วยในการประมวลผลข้อมูล โดยนำความสัมพันธ์ของเพจที่ได้จากขั้นตอนการเก็บข้อมูลมาหาระยะทางระหว่างเพจ และจัดกลุ่มด้วยการแบ่งคลัสเตอร์แบบเค-มีนส์ (K-Means Clustering) ผลลัพธ์เพื่อต้องการให้เพจที่มีความคล้ายกันอยู่กลุ่มเดียวกันเพื่อใช้ในการแนะนำเพจที่มีความคล้ายกับความสนใจของผู้ใช้งาน

1.4.3 การสรุปผลการทดลอง

ขั้นตอนนี้คือการนำข้อมูลที่ได้จากการเก็บข้อมูลและประมวลผลข้อมูลมาสรุปผลการทดลอง เปรียบเทียบกับวิธีต่างๆ และนำเสนอข้อมูลทั้งหมดที่เก็บมาได้ และมีลักษณะเป็นอย่างไร

1.5 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อแนะนำแฟนเพจที่ตรงกับความสนใจของผู้ใช้งานเฟซบุ๊ก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

ในการหาความสัมพันธ์ระหว่างเฟซบุ๊กเพจนั้นต้องใช้ทฤษฎีที่เกี่ยวข้องในหลายๆด้านเข้ามาช่วยในการหาคำตอบ โดยจะแยกกลุ่มของทฤษฎีที่เกี่ยวข้องออกเป็น 2 ส่วนด้วยกันคือ วิธีการทำคลัสเตอร์ (clustering) และ การวิเคราะห์โซเชียลเน็ตเวิร์ก (social network analysis)

2.1 การทำคลัสเตอร์ (Clustering)

การทำคลัสเตอร์เป็นวิธีการวิเคราะห์ข้อมูล ใช้ในการเรียนรู้ของเครื่อง โดยการแบ่งข้อมูลออกเป็นกลุ่ม โดยข้อมูลที่มีคุณลักษณะคล้ายกันจะอยู่ด้วยกัน แต่มีความแตกต่างกับกลุ่มอื่นๆ โดยคุณลักษณะที่เลือกใช้จะต้องมีความแตกต่างกันกับคุณลักษณะอื่นๆอย่างชัดเจน มิเช่นนั้นจะเป็นการแบ่งแยกที่ไม่ชัดเจน และมีความผิดพลาดได้ ตัวอย่างของการเลือกคุณสมบัติในการแบ่งกลุ่ม เช่น การแบ่งกลุ่มบ้านโดยพิจารณาคุณลักษณะต่างๆของบ้านเพื่อเป็นเครื่องมือในการตัดสินใจในการเลือกซื้อ ในที่นี่เลือกคุณลักษณะที่ใช้พิจารณาคือ ราคา และจำนวนห้อง

การทำคลัสเตอร์เป็นการแบ่งกลุ่มข้อมูลโดยไม่มีการกำหนดประเภทของข้อมูลไว้ก่อน จึงกล่าวได้ว่าการแบ่งกลุ่มข้อมูล เป็นการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) ซึ่งแตกต่างจากการจำแนกประเภทข้อมูล (classification) ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (supervised learning)

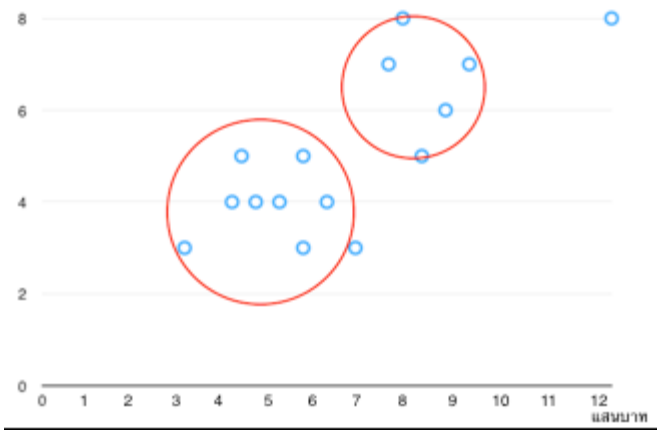
Chulalongkorn University

ตารางที่ 2 แสดงคุณลักษณะของบ้าน

บ้าน	ราคา (บาท)	จำนวนห้อง
A	400,000	4
B	300,000	3
C	500,000	4
D	600,000	4
E	550,000	5
F	420,000	5
G	450,000	4
H	800,000	5
I	850,000	6
J	900,000	7
K	730,000	7
L	760,000	8
M	1,200,000	8
N	66,000	3
O	550,000	3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 2 เห็นได้ว่าบ้านแต่ละหลังคุณลักษณะอยู่ 2 คุณลักษณะคือ ราคา และจำนวนห้อง ซึ่งหากนำมาทำเป็นกราฟและใช้ความชำนาญของผู้ใช้ จะมองเห็นกลุ่มแยกเป็น 2 กลุ่มอย่างชัดเจนดังแสดงในรูปที่ 5 แต่หากว่าข้อมูลมีความซับซ้อนมากยิ่งขึ้น ก็อาจจะมองเห็นกลุ่มได้ยากมากขึ้น จึงต้องมีวิธีที่ช่วยในการแบ่งกลุ่มที่มีประสิทธิภาพมากยิ่งขึ้น



รูปที่ 5 แสดงการแบ่งกลุ่มของการแบ่งคุณลักษณะบ้าน

2.1.1 วัตถุประสงค์ของการทำคลัสเตอร์

เมื่อทำคลัสเตอร์เสร็จเรียบร้อยแล้ว ผลลัพธ์คือ จุดตัวอย่างที่มีคุณลักษณะคล้ายกันจะอยู่กลุ่มเดียวกัน ทำให้สามารถแยกแยะความแตกต่าง และสามารถตอบสนองต่อกลุ่มที่แตกต่างกันด้วยการกระทำที่แตกต่างกัน อีกทั้งยังสามารถลดขนาดของข้อมูลได้อีกด้วย เมื่อเราไม่จำเป็นต้องเก็บข้อมูลทั้งหมดเอาไว้เพื่อพิจารณาว่า จุดตัวอย่างใหม่ที่เข้ามาจะอยู่ในกลุ่มใด สามารถใช้กลุ่มที่จัดเอาไว้แล้วแยกแยะจุดตัวอย่างที่เข้ามาใหม่ได้ทันที เช่นการวิเคราะห์แบ่งกลุ่มคุณลักษณะของบ้านเพื่อการเลือกซื้อจากตัวอย่างข้างต้น จะทำให้เห็นความคุ้มค่าในการซื้อบ้านว่าเลือกซื้อบ้านแบบใดดีถึงจะคุ้มค่ามากที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.2 การแบ่งคลัสเตอร์แบบเค-มีนส์ (K-Means Clustering)

เค-มีนส์ คือหนึ่งในวิธีการเรียนรู้แบบไม่มีผู้สอนที่ง่ายที่สุดวิธีหนึ่งที่สามารถแก้ไขปัญหาที่รู้จักกันดีในเรื่องการแบ่งกลุ่มข้อมูล ขั้นตอนการแบ่งกลุ่มจะแบ่งตามข้อมูลตั้งต้นว่าจะให้แบ่งกลุ่มข้อมูลทั้งหมดออกเป็นกี่กลุ่ม แนวคิดหลักของวิธีการนี้คือการวางตำแหน่งของ k (จุดศูนย์กลางของแต่ละคลัสเตอร์) ซึ่งการวางตำแหน่งเริ่มต้นของ k นั้นมีผลต่อผลลัพธ์เป็นอย่างมาก จึงต้องวิธีในการวางที่ดี วิธีที่ง่ายที่สุดคือการวางตำแหน่งเริ่มต้นทั้ง k จุดให้ห่างกันที่สุด ไม่ควรเลือกตำแหน่งเริ่มต้นให้ใกล้กันจนเกินไป เพราะจะได้กลุ่มที่มีความคล้ายคลึงกันมากเกินไป ลำดับถัดไปคือหาระยะทางจากจุดต่างๆว่าอยู่ใกล้กับ k มากเพียงใด ให้ k ที่ใกล้กับแต่ละจุดที่สุดเป็นตัวแทนของจุดต่างๆเหล่านั้น ในรอบแรกก็จะได้คลัสเตอร์แล้ว แต่เราต้องคำนวณใหม่เพื่อหาคลัสเตอร์ในรอบถัดไป โดยการคำนวณหาตำแหน่งของ k เซ็นทรอยด์ (centroid) ใหม่จากค่าเฉลี่ยตำแหน่งของจุดต่างๆในคลัสเตอร์นั้นๆของรอบที่แล้ว จากนั้นจึงทำการจัดกลุ่มใหม่โดยดูจากระยะห่างของจุดต่างๆกับเซ็นทรอยด์ กระบวนการ

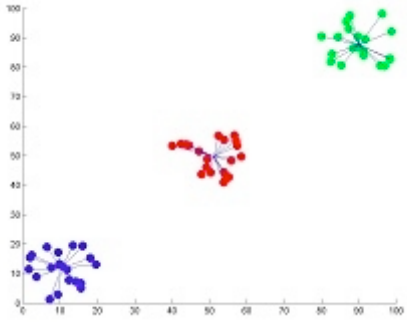
นี้จะจบเมื่อตำแหน่งของเซ็นทรอยด์ และสมาชิกในคลัสเตอร์ไม่มีการเปลี่ยนแปลงแล้ว หรือเปลี่ยนแปลงน้อยที่สุด

ขั้นตอนของการแบ่งคลัสเตอร์แบบเค-มีนส์ มีดังนี้

1. เลือก เซ็นทรอยด์ c จำนวน k ตัวแบบสุ่ม
2. คำนวณหาระยะห่างระหว่างตำแหน่งของข้อมูลกับ c
3. ให้จัดคลัสเตอร์สำหรับข้อมูลทุกตัว โดยจัดให้ข้อมูลอยู่ในคลัสเตอร์ที่มีเซ็นทรอยด์ใกล้ที่สุด
4. คำนวณหาตำแหน่งของ c ในแต่ละคลัสเตอร์ใหม่
5. คำนวณระยะห่างระหว่างตำแหน่งของข้อมูลกับ c ใหม่
6. ถ้าสมาชิกของแต่ละ cluster ไม่เปลี่ยนแปลง หรือเปลี่ยนแปลงน้อยแล้วให้หยุด แต่ถ้าไม่ใช่ ให้ไปทำกระบวนการของข้อ 3-6 จนกว่าสมาชิกในคลัสเตอร์ไม่เปลี่ยนแปลง หรือเปลี่ยนแปลงน้อยกว่าค่าที่กำหนดไว้

ข้อดีของการแบ่งคลัสเตอร์แบบเค-มีนส์

1. เร็ว และง่ายต่อการทำความเข้าใจ
2. ถ้าเปรียบเทียบประสิทธิภาพแล้ว จะพบว่า จำนวนคลัสเตอร์, จำนวนมิติของข้อมูล และจำนวนรอบของการคำนวณ จะน้อยกว่าจำนวนตำแหน่งของข้อมูลเป็นอย่างมาก
3. มีประสิทธิภาพมากในกรณีที่ข้อมูลมีการกระจายตัวแบบแยกออกจากกันเป็นกลุ่มๆ ดังแสดงในรูปที่ 6



รูปที่ 6 แสดงการกระจายตัวของข้อมูล

ข้อเสียของการแบ่งคลัสเตอร์แบบเค-มีนส์

1. การเรียนรู้แบบนี้ต้องการตำแหน่งเริ่มต้นของจุดศูนย์กลางคลัสเตอร์ซึ่งผลลัพธ์ และประสิทธิภาพจะขึ้นอยู่กับตำแหน่งเริ่มต้นนี้

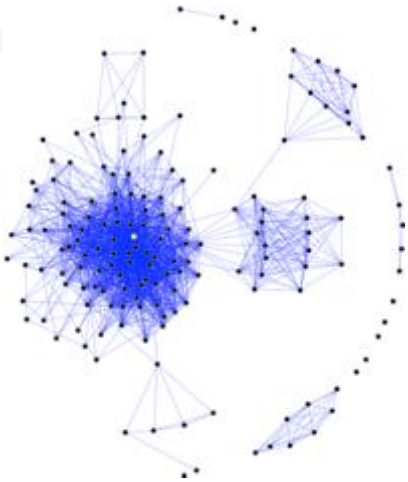
2. หาค่า k ที่เหมาะสมได้ยาก

3. อ่อนไหวต่อข้อมูลรบกวน (noise)

2.2 การวิเคราะห์โซเชียลเน็ตเวิร์ก

การวิเคราะห์โซเชียลเน็ตเวิร์ก คือการวิเคราะห์เครือข่าวความสัมพันธ์ทางสังคมซึ่งประกอบไปด้วยโหนด (nodes) และไทส์หรือการเชื่อม (ties) ตัวอย่างเช่น เส้นเชื่อม (edge), ลิงก์ (link) และการเชื่อมต่อ (connection) โดยที่โหนดหมายถึงตัวตนหนึ่งๆในเครือข่ายทางสังคมเช่น บุคคล บริษัท หรือเมือง และไทส์ หมายถึงความสัมพันธ์ระหว่างโหนด เช่นถ้าโหนดคือเมือง ไทส์จะเป็นถนน ถ้าโหนดคือบุคคล ไทส์อาจเป็นความสัมพันธ์แบบเพื่อน

ผลลัพธ์ของรูปแบบความสัมพันธ์ค่อนข้างมีความซับซ้อน อาจเป็นได้หลายๆรูปแบบของไทส์ (เพื่อน ครอบครัว) ระหว่างโหนด (บุคคล) เป็นได้ถึงรูปแบบความสัมพันธ์ในระดับครอบครัว ไปจนถึงความสัมพันธ์ในระดับประเทศชาติ แต่ถ้ามองในมุมมองที่ง่ายขึ้น อาจมองในรูปแบบของการเชื่อมต่อกันในรูปแบบที่เฉพาะเจาะจงของไทส์ กับโหนดที่สนใจศึกษาเท่านั้น ซึ่งโครงสร้างความสัมพันธ์ที่เกิดขึ้นจะสามารถวัดผลได้ดีมาก และมักจะแสดงผลออกมาในรูปแบบของกราฟความสัมพันธ์ที่โหนดใดๆจะแทนเป็นจุด และไทส์ที่เชื่อมระหว่างโหนดจะแสดงเป็นเส้น ดังรูปที่ 7



รูปที่ 7 แสดงความสัมพันธ์ระหว่าโหนดโดยมีไทส์เป็นเส้นเชื่อม

2.2.1 การวัดผลของการวิเคราะห์โซเชียลเน็ตเวิร์ก

สามารถมองได้ในหลายมุมมองได้แก่

2.2.1.1 สะพาน (Bridge)

เป็นไทส์ที่เชื่อมโยงระหว่างกลุ่มของโซเชียลเน็ตเวิร์ก 2 กลุ่ม หากตัดการเชื่อมโยงเส้นนี้ไปแล้ว จะทำให้โซเชียลเน็ตเวิร์ก 2 กลุ่มนี้แยกออกจากกันชัดเจน ไม่มีความสัมพันธ์ใดที่เชื่อมถึงกันเลย

2.2.1.2 ผู้เป็นสะพาน (Bridger)

เป็นโหนดหรือคนที่มีการเชื่อมโยงไปยังกลุ่มของโซเชียลเน็ตเวิร์กกลุ่มอื่นๆ ซึ่งโหนดใดที่เป็นผู้เป็นสะพานนั้นจะสร้างผลกระทบต่อกลุ่มที่มีการเชื่อมโยงกันอยู่ สามารถคำนวณหาผู้เป็นสะพานได้จากการหาค่า ความเป็นจุดศูนย์กลางระหว่างกัน (betweenness centrality) และ ข้อจำกัดของเน็ตเวิร์ก (network constraint) โดยที่โหนดใดที่มีค่าความเป็นจุดศูนย์กลางระหว่างกันสูง และมีค่าการจำกัดของเน็ตเวิร์กต่ำ จะเป็นผู้เป็นสะพาน

2.2.1.3 ฮับ (Hubs)

คือโหนดที่มีอิทธิพลมากที่สุดในเครือข่าย โดยใช้ค่า ความเป็นจุดศูนย์กลางแบบไม่มีระดับ (indegree centrality) ในการคำนวณ ซึ่งนับได้จากจำนวนลิงก์แบบมีทิศทางที่เข้ามาหาโหนดนั้น

2.2.1.4 ความเป็นจุดศูนย์กลาง (Centrality)

คือค่าที่ใช้วัดความเป็นจุดศูนย์กลางของเครือข่ายว่ามีความเป็นจุดศูนย์กลางที่ดีเพียงไร โดยสามารถแบ่งการพิจารณาความเป็นจุดศูนย์กลางออกเป็น 4 แบบคือ

1. ความเป็นจุดศูนย์กลางแบบดีกรี (Degree Centrality)

เป็นการค้นหาว่าสมาชิกใดเป็นจุดศูนย์กลางโดยดูว่าใครเป็นฮับของเครือข่าย ซึ่งเป็นตำแหน่งที่มีอิทธิพลมากที่สุดในกลุ่มนั้นๆ โดยวัดจากลิงก์แบบมีทิศทางที่เข้ามาสู่โหนดนั้นๆ

2. ความเป็นจุดศูนย์กลางแบบใกล้ชิด (Closeness Centrality)

เป็นการค้นหาว่าสมาชิกใดเป็นจุดศูนย์กลางโดยดูจากความใกล้ชิดกับสมาชิกอื่นๆ และใช้ระยะทางที่สั้นที่สุดในการเข้าถึง โดยการวัดจากจำนวนเส้นเชื่อมโยงทั้งหมดที่ใช้ในการเดินทางจากสมาชิกหนึ่งไปยังอีกสมาชิกหนึ่ง โดยการลากผ่านสมาชิกตัวอื่นๆด้วยระยะทางที่สั้นที่สุด

3. ความเป็นจุดศูนย์กลางระหว่างกัน (Betweenness Centrality)

เป็นการค้นหาว่าสมาชิกใดบ้างที่มีตำแหน่งเป็นสะพานที่เชื่อมกลุ่มต่างๆเข้าหากัน โดยใช้วิธีการคำนวณจากสัดส่วนของระยะทางที่สั้นที่สุดในการเชื่อมโยงสมาชิกแต่ละคู่

4. ความเป็นจุดศูนย์กลางของเวคเตอร์ลักษณะเฉพาะ (Eigenvector Centrality)

เป็นการวัดค่าอิทธิพลของสมาชิกเครือข่ายโดยใช้หลักการคือสมาชิกที่เชื่อมโยงกับสมาชิกอื่นที่มีค่าอิทธิพลสูงจะมีค่าเวคเตอร์ลักษณะเฉพาะที่สูง เมื่อเทียบกับการเชื่อมโยงกับสมาชิกที่มีค่าอิทธิพลต่ำ กรณีนี้ยกตัวอย่างได้เป็นอย่างดีเช่น กูเกิลเพจแรงก์ (Google PageRank)

2.2.1.5 ความหนาแน่น (Density)

คือค่าที่ใช้วัดความหนาแน่นของเครือข่ายโดยวัดจาก จำนวนไทส์ที่มีอยู่จริงในเครือข่าย หารด้วยจำนวนไทส์ที่เป็นไปได้ทั้งหมด

2.2.1.6 ลิงก์ต่อโหนด (Links Per Node)

คือจำนวนลิงก์ต่อโหนดโดยวัดจาก จำนวนไทส์ที่มีอยู่จริงในเครือข่าย หารด้วยจำนวนโหนดทั้งหมดที่มีอยู่ในเครือข่าย

2.2.1.7 ความสมดุลของโครงสร้าง (Structural Equivalence)

คือความสัมพันธ์ของโหนดที่มีการเชื่อมโยงไปยังโหนดอื่นๆร่วมกัน โดยที่ระหว่างโหนดนั้นๆไม่จำเป็นต้องมีความสัมพันธ์กันเอง ยกตัวอย่างเช่น คนที่ซื้อหนังสือ A และ B ไปแล้ว มักมีแนวโน้มที่จะซื้อหนังสือ C และ D

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.3 งานวิจัยที่เกี่ยวข้อง

เฟซบุ๊กเป็นโซเชียลเน็ตเวิร์กที่ใหญ่ที่สุดในโลก และมีผู้ใช้งานกว่า 1 พันล้านคน (ตัวเลขจากปี พ.ศ.2558) ซึ่งพฤติกรรมของผู้ใช้งานเฟซบุ๊กจะแตกต่างกับผู้ใช้งานเว็บไซต์ทั่วไป โดยที่งานวิจัย [1] กล่าวไว้ว่าการใช้งานเฟซบุ๊กในแต่ละเพศ วัย และประสบการณ์ที่แตกต่างกันไปของแต่ละบุคคล ย่อมส่งผลถึงพฤติกรรมการใช้งานที่แตกต่างกัน และหากมองในเรื่องการโฆษณา หรือการทำการตลาดกับผู้ใช้งาน ก็ย่อมมีผลที่แตกต่างกันเช่นกัน ทำให้เราไม่สามารถนำรูปแบบของการใช้งานเว็บไซต์มาใช้ในการแบ่งแยกกลุ่มผู้ใช้งานในเฟซบุ๊กได้

มีความพยายามแบ่งกลุ่มผู้ใช้งานเฟซบุ๊กที่เข้ามาใช้งานแฟนเพจออกเป็นกลุ่มๆ เพื่อหากลุ่มลูกค้าที่น่าจะมีกำลังซื้อ ซึ่งเป็นการประหยัดต้นทุนในการทำการตลาดได้เป็นอย่างมาก และสามารถแบ่งกลุ่มผู้ใช้งานออกมาได้เพียงทำการวิเคราะห์เฟซบุ๊กโพสต์ และคอมเมนต์ด้วยวิธีการทำเหมือง

ข้อมูล และการประมวลผลภาษาธรรมชาติ [2] แต่จริงๆแล้วจำนวนผู้ใช้งานเฟซบุ๊กที่มีคอมเมนต์ยังมีอยู่น้อย จึงทำให้ไม่สามารถวิเคราะห์ข้อมูลออกมาได้ดีพอ

หากต้องการเก็บข้อมูลการใช้งานด้วยวิธีการใช้เฟซบุ๊กแอพพลิเคชั่นคือการเขียนแอพพลิเคชั่นเพื่อใช้งานบนเฟซบุ๊กนั้น จำเป็นต้องขออนุญาตจากผู้ใช้งานในการเข้าถึงข้อมูล ในกรณีที่ผู้ใช้งานเข้าใจความเสี่ยงในการอนุญาตให้ผู้พัฒนาแอพพลิเคชั่นเข้าถึงข้อมูลดังกล่าวแล้ว แต่ก็อาจมีผู้ใช้งานบางส่วนที่ไม่ยอมให้ข้อมูล การที่ผู้ใช้งานจะอนุญาตให้เข้าถึงข้อมูลได้นั้น ขึ้นอยู่กับผลประโยชน์ที่จะได้รับและความภักดีต่อแบรนด์นั้นๆ [3] วิธีการเก็บข้อมูลแบบนี้จึงเป็นไปได้ยากในกรณีที่เป็นแบรนด์หน้าใหม่ หรือไม่มีการนำเสนออะไรที่น่าสนใจกับผู้ใช้งานจริงๆ และด้วยการเข้าถึงข้อมูลต่างๆของผู้ใช้งานได้แล้ว เรายังสามารถรู้ได้ถึงความเห็นด้านบวก หรือลบได้อีกด้วยจากการค้นหาคำที่ใช้บ่อยในทวิตเตอร์ (Twitter) ว่าอยู่ในกลุ่มความคิดเห็นด้านบวกหรือลบ เพื่อไปใช้คัดแยกความคิดเห็นในเฟซบุ๊ก [4]

งานวิจัย [5] ที่ศึกษาอิทธิพลของคนที่กดชื่นชอบแฟนเพจ ภาพ หรือข้อความต่างๆของแฟนเพจ มีผลต่อการกำลังซื้อของคน ยิ่งถ้าเป็นเพื่อนกันก็จะยิ่งมีผลต่อโอกาสในการซื้อของ สำหรับผู้ใช้งานเฟซบุ๊กในประเทศไทย

จากการศึกษาโซเชียลเน็ตเวิร์กของประเทศจีนชื่อว่า ซิน่า (Sina) พบว่าผู้ใช้งานที่มีความสัมพันธ์กัน มักจะมีความสนใจอะไรที่คล้ายๆกัน [6] แม้แต่ในการศึกษาเกี่ยวกับโซเชียลเน็ตเวิร์กชื่อทวิตเตอร์ ได้มีการวิเคราะห์ความสัมพันธ์ของผู้ติดตามซึ่งเป็นความสัมพันธ์แบบทิศทางเดียว ยังพบว่ามีความสัมพันธ์กันที่น่าสนใจ [7] เช่น ความสัมพันธ์ระหว่างผู้ใช้งานแต่ละคนแบบเดียวกันแบบเพจแรงค์ (PageRank), การกระจายตัวของทวิตเตอร์ไม่ได้หยุดแค่ 1 ฮ็อบ (hop) แต่ไปถึง 3-4 ฮ็อบ

จากความสัมพันธ์ต่างๆที่เกิดขึ้นของโซเชียลเน็ตเวิร์กในงานวิจัยต่างๆ สามารถวิเคราะห์ออกมาในรูปแบบของการวิเคราะห์โซเชียลเน็ตเวิร์ก (social network analysis) ได้ [8] จึงสามารถมองโซเชียลเน็ตเวิร์กออกมาเป็นหลายมิติมากยิ่งขึ้น และเข้าใจได้ง่ายขึ้น

เนื่องจากผู้ใช้งานในโซเชียลเน็ตเวิร์กมีจำนวนมาก การจะแยกแยะว่าผู้ใช้งานคนไหนเป็นคนที่ชอบอะไร หรืออยู่กลุ่มไหนย่อมเป็นไปได้จาก จึงมีงานวิจัย [9] ที่พยายามทำคลัสเตอร์ของผู้ใช้งานในการจำแนกผู้ใช้งานออกเป็นกลุ่ม ทำให้ง่ายต่อการมองภาพรวมและทำความเข้าใจโซเชียลเน็ตเวิร์กได้ดีขึ้นมาก

ดังนั้นการค้นหาความสัมพันธ์ของเฟซบุ๊กแฟนเพจที่เปรียบเสมือนการมองแฟนเพจเป็นบุคคลย่อมมีความเป็นไปได้มากขึ้น หากมีการทำคลัสเตอร์ และวิเคราะห์ข้อมูลด้วยการวิเคราะห์โซเชียลเน็ตเวิร์

บทที่ 3
แนวคิดและวิธีการดำเนินงาน

ตามปกติแล้วทุกสิ่งทุกอย่างมีความสัมพันธ์เชื่อมโยงกันเสมอ ไม่ว่าจะเป็นความสัมพันธ์แบบทิศทางเดียว สองทิศทาง หรือไม่มีทิศทาง เช่นความสัมพันธ์ระหว่างเมือง เชื่อมโยงกันด้วยถนนที่ทอดผ่านไป หรือเป็นความสัมพันธ์ของเว็บไซต์ ก็มีการเชื่อมโยงของลิงก์ระหว่างกันแบบมีทิศทาง หากมองในเรื่องความสัมพันธ์ของคน ก็จะใช้คำว่าเพื่อนเป็นตัวเชื่อมโยงแบบไม่มีทิศทางระหว่างคน หากสามารถหาความสัมพันธ์ของเครือข่ายโซเชียลเน็ตเวิร์กที่ใหญ่ที่สุดในโลกอย่างเฟซบุ๊กได้ ก็จะยิ่งมีความน่าสนใจมากยิ่งขึ้น

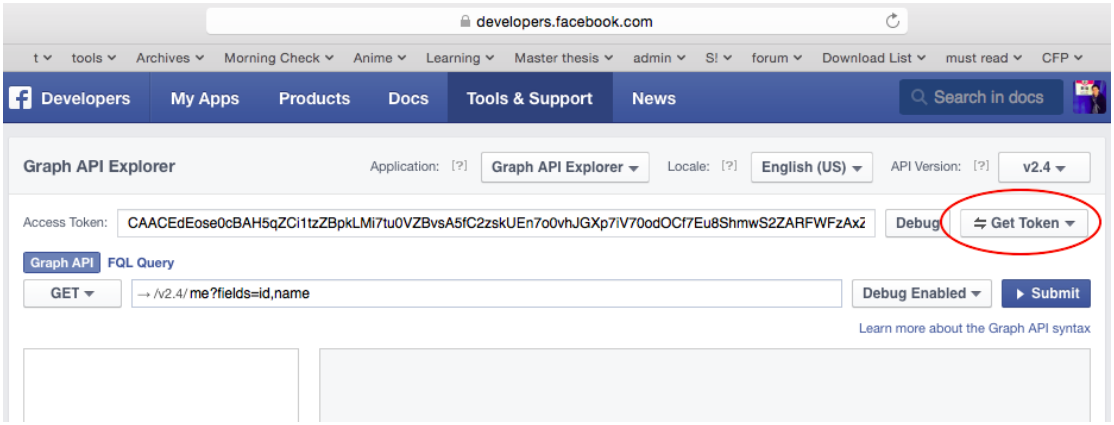
เพจเป็นเครื่องมือของเฟซบุ๊กที่ทำให้ผู้ใช้งานสามารถแบ่งปันเรื่องราวต่างๆได้เสมือนหนึ่งว่ามีเว็บไซต์เป็นของตนเอง ผู้ใช้งานจะกดไลค์เพื่อติดตามแฟนเพจที่ตรงกับความสนใจกับตนเอง โดยที่ผู้ใช้งานจะกดไลค์เพจต่างๆได้จากการแนะนำของเพื่อน เช่นเห็นข้อความต่างๆที่แชร์มาจากเพื่อนแล้วเกิดความสนใจ หรือจากเพจที่น่าสนใจอยู่ในกระแสของสังคม แต่ผู้ใช้งานไม่อาจจะรู้ได้เลยว่ามีเพจที่น่าสนใจอื่นๆอีกที่น่าจะตรงกับความสนใจของตนเอง

ด้วยปัญหานี้จึงเกิดแนวคิดที่ใช้แก้ปัญหาเพื่อจะสามารถแนะนำเพจที่ตรงกับความสนใจของผู้ใช้งานได้ แนวคิดและวิธีการดำเนินงานเพื่อแนะนำเพจที่ตรงกับความสนใจของผู้ใช้งานมีดังต่อไปนี้

3.1 หาผู้ใช้งานตัวอย่าง

งานวิจัยนี้เริ่มต้นด้วยการหาผู้ใช้งานตัวอย่างเพื่อใช้ในการเก็บข้อมูลเป็นลำดับ โดยผู้ใช้งานตัวอย่างนี้หามาจากผู้ใช้งานของเพจธนาคารที่มีคนติดตามมากที่สุดของไทยคือ SCB Thailand (https://www.facebook.com/scb.thailand) มีคนติดตาม 2.9 ล้านคน และ KBank Live (https://www.facebook.com/KBankLive) มีคนติดตาม 1.8 ล้านคน โดยการเก็บข้อมูลผู้กดไลค์และผู้กดคอมเมนต์ที่ทำกับโพสต์ของเพจนี้

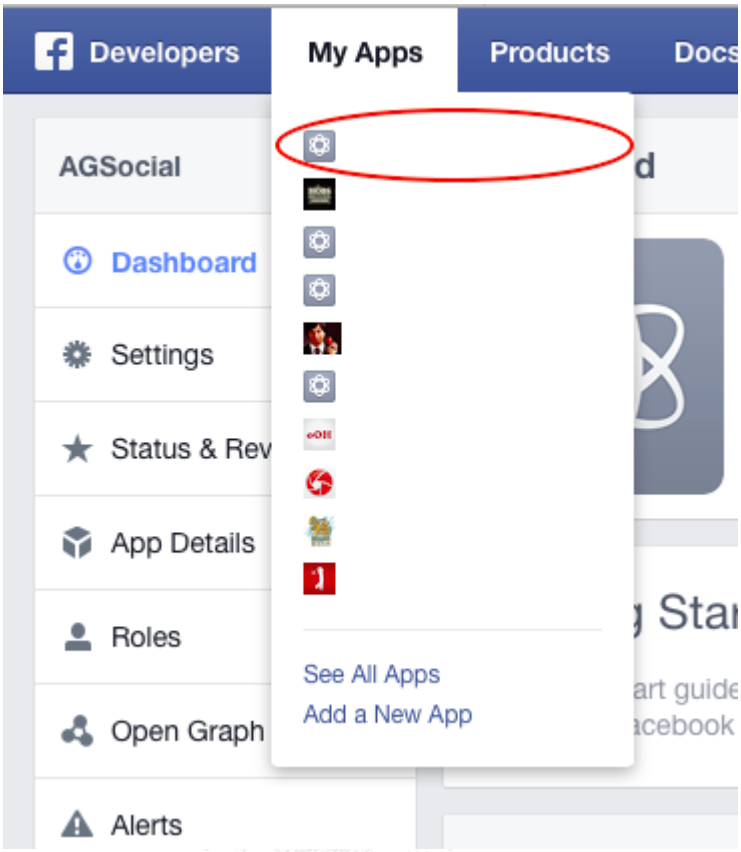
การเก็บข้อมูลนี้ต้องใช้กราฟเอพีไอในการเรียกข้อมูลจากเฟซบุ๊ก แต่ก่อนที่จะเรียกใช้กราฟเอพีไอได้นั้น ต้องขอโทเคนสำหรับการเข้าถึง (Access Token) ก่อนเพื่อใช้งานข้อมูล เพราะเฟซบุ๊กค่อนข้างจะมีข้อกำหนดเรื่องความเป็นส่วนตัวที่มากขึ้นเรื่อยๆ ไม่ใช่ว่าเป็นผู้ใช้งานคนไหนก็ได้ที่จะเรียกข้อมูลได้ จึงต้องขอโทเคนก่อน โดยการขอโทเคนสามารถทำได้โดยเข้าไปที่ https://developers.facebook.com/tools/explorer/ เพื่อขอโทเคน (จำเป็นต้องสร้างแอพพลิเคชั่นกับเฟซบุ๊กก่อน จากนั้นจึงสามารถขอโทเคนได้)



รูปที่ 8 แสดงการเข้าไปขอโทเคน

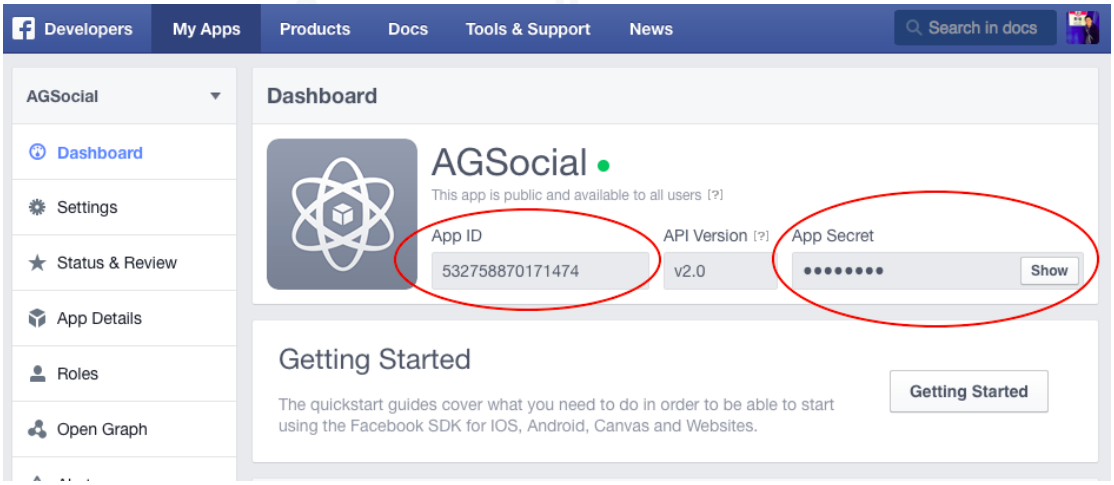
จากรูปที่ 8 จะเห็นปุ่ม “Get Token” เพื่อใช้ในการขอโทเคน ซึ่งทุกครั้งที่มีการเรียกใช้กราฟเอพีไอจะต้องแนบโทเคนเป็นตัวแปรในลิงก์ด้วยเสมอ แต่โทเคนที่ขอได้ในขั้นตอนนี้อยู่ในระดับโทเคนของผู้ใช้งาน ซึ่งมีอายุการใช้งานเพียงแค่ 2 ชั่วโมงเท่านั้น และมีการใช้งานที่ค่อนข้างจำกัด สามารถเรียกใช้งานกราฟเอพีไอด้วยโทเคนได้ในจำนวนที่ค่อนข้างน้อยภายในช่วงเวลาหนึ่ง เมื่อถึงจำนวนที่จำกัดแล้วจะไม่สามารถเรียกใช้งานได้ 30 นาที ซึ่งในการเก็บข้อมูลในลักษณะนี้จะไม่สะดวกเป็นอย่างมาก

เฟซบุ๊กจำกัดสิทธิ์ในการเข้าถึงข้อมูลด้วยระดับของโทเคนจึงต้องทำการขอโทเคนอีกระดับเพื่อระยะเวลาในการใช้งานที่นานขึ้น และมีจำนวนครั้งในการเรียกใช้งานที่มากขึ้น โดยการขอในระดับแอพพลิเคชั่นโทเคนซึ่งผู้ที่จะขอโทเคนระดับนี้ได้ต้องเป็นผู้ที่เป็นเจ้าของแอพพลิเคชั่นเพราะต้องใช้ app-id และ app-secret โดยดูข้อมูลได้จากการคลิกที่ “My Apps” เพื่อดูแอพพลิเคชั่นที่ต้องการขอโทเคนดังรูปที่ 9



รูปที่ 9 แสดงแอพพลิเคชั่นที่ผู้ใช้งานเป็นเจ้าของ

จากนั้นจึงสามารถเห็นข้อมูลของ app-id และ app-secret ดังรูปที่ 10



รูปที่ 10 แสดง App ID และ App Secret ของแอพพลิเคชั่นที่เลือก

เมื่อได้ app-id และ app-secret แล้วจึงนำไปใช้เพื่อขอแอพพลิเคชั่นโทเคนได้จากการเรียกใช้งาน https://graph.facebook.com/oauth/access_token?client_id={app-id}&client_secret={app-secret}&grant_type=client_credentials

เมื่อได้โทเคนที่ต้องการแล้ว จึงสามารถเรียกใช้งานกราฟเอพีไอได้โดยข้อมูลที่ต้องการเรียกใช้งานในครั้งนี้คือผู้กดไลค์ และผู้กดคอมเมนต์ของโพสต์ ก่อนอื่นให้เรียกข้อมูลโพสต์ของเพจก่อนด้วย https://graph.facebook.com/v2.0/[page_id]?field=posts&access_token=[access-token]


```
{
  "posts": {
    "data": [
      {
        "message": "มาแล้ว!!! รายละเอียดการเปลี่ยนแปลงเงื่อนไขการลงทุน LTF และ RMF แบบเข้าใจง่าย พร้อมให้ทุกคนได้ทำความเข้าใจ เพื่อให้วางแผนลงทุนกันได้อย่างถูกต้อง พร้อมแล้วก็คลิกไปอ่านต่อกันได้เลยค่ะ

ทุกเรื่องการเงิน "K-Expert ตอบได้ ช่วยได้" ที่ www.askkbank.com/k-expert และปรึกษาได้ฟรีทางอีเมล k-expert@kasikornbank.com",
        "created_time": "2015-07-13T09:05:10+0000",
        "id": "127610057890_10153274431887891"
      },
      {
        "message": "ลูกค้ากองทุนกสิกรไทย ยิ้มรับเงินปันผลจากกองทุนหุ้นไทย 2 กองทุน ลงทุนได้ง่ายๆกับหลากหลายกองทุนที่คัดสรรมาเพื่อคุณ! คลิกอ่านต่อได้ ที่นี่เลยค่ะ http://goo.gl/Hq7AmU",
        "created_time": "2015-07-13T07:30:36+0000",
        "id": "127610057890_10153274358882891"
      },
      {
        "message": "หากใครเดินทางไปเที่ยวหรือไปเรียนต่อที่สหราชอาณาจักร ( UK ) แล้วต้องการกดเงิน ถ้าเห็นเขาเขียนว่า " Hole in the Wall " ก็อย่าได้งงกันไปนะ เพราะมันก็คือ " ตู้ ATM " นั่นเอง

ทุกเรื่องการเงิน " K-Expert ตอบได้ ช่วยได้ " ที่ www.askkbank.com/k-expert และปรึกษาได้ฟรีทางอีเมล k-expert@kasikornbank.com",
        "created_time": "2015-07-12T03:00:00+0000",
        "id": "127610057890_10153265849537891"
      },
      {
        "message": "เพื่อความปลอดภัยในการใช้งาน Online Banking ขอเน้นย้ำกับเพื่อนๆว่า การใช้บริการ Online ทางธนาคารจะไม่มีการขอข้อมูลส่วนตัวใดๆเพิ่มเติม โดยหากลูกค้าต้องการทราบความเคลื่อนไหวหรือข่าวของธนาคาร สามารถติดตามได้ที่เว็ปไซต์ www.kasikornbank.comหรือ Facebook KBank Live เท่านั้นนะคะ",
        "created_time": "2015-07-11T09:00:01+0000",
        "id": "127610057890_10153267815057891"
      },
```


รูปที่ 11 แสดงข้อมูลโพสต์ที่เรียกได้จากกราฟเอพีไอ

จากรูปที่ 11 เห็นได้ว่าแต่ละโพสต์จะมี post_id ให้เก็บไว้เพื่อใช้ในการเรียกข้อมูลของผู้กดไลค์ และ ผู้กดคอมเมนต์จาก https://graph.facebook.com/v2.0/[post_id]?access_token=[access-token] ซึ่งได้ผลลัพธ์ตามรูปที่ 12

รูปที่ 12 แสดงข้อมูลที่เรียกได้จากโพสต์

จากรูปที่ 12 จะแสดงข้อมูลที่เรียกได้จากโพสต์ ซึ่งมีข้อมูลของผู้กดไลค์ และผู้กดคอมเมนต์ ซึ่งสามารถใช้ข้อมูลส่วนนี้ในการหาผู้ใช้งานตัวอย่างได้

จำนวนของผู้ใช้งานตัวอย่างใช้เพียง 400 คน ตามวิธีคำนวณกลุ่มประชากรตัวอย่างของ Taro Yamane [10] เพื่อเป็นตัวแทนของผู้ใช้งานเฟซบุ๊กจำนวน 1.4 พันล้านคน ด้วยความผิดพลาดร้อยละ 5

3.2 หาเพจที่ผู้ใช้งานตัวอย่างติดตาม

เพื่อค้นหาความสัมพันธ์ที่เกิดขึ้นจากผู้ใช้งานตัวอย่าง จึงต้องหาว่าผู้ใช้งานตัวอย่างได้ติดตามเพจใดบ้าง การหาข้อมูลนี้สามารถหาได้จากกราฟเอพีไอ แต่จำเป็นต้องได้โทเคนของผู้ใช้งานที่ต้องการเก็บข้อมูล จึงเป็นไปไม่ได้เลยที่จะได้โทเคนนี้มาโดยง่าย มีอีกวิธีหนึ่งที่สามารถหาข้อมูลเพจที่ผู้ใช้งานติดตามโดยไม่จำเป็นต้องใช้กราฟเอพีไอนั่นคือการที่เข้าไปยังหน้าโปรไฟล์ของผู้ใช้งานโดยตรงในขณะที่ยังไม่ได้เข้าสู่ระบบ โดยเข้าไปที่ http://facebook.com/[user_id]

รูปที่ 13 แสดงข้อมูลเพจที่ผู้ใช้งานติดตามจากหน้าโปรไฟล์

จากรูปที่ 13 หากผู้ใช้งานเปิดการเข้าถึงข้อมูลของเพจที่ติดตามอยู่เป็นสาธารณะ จะทำให้สามารถดึงข้อมูลส่วนนี้มาใช้งานได้ แต่หากผู้ใช้งานคนใดที่ไม่ได้เปิดให้เป็นสาธารณะ ก็จะไม่สามารถดึงข้อมูลได้ ข้อมูลนี้อยู่ในลักษณะของเฮชทีเอ็มแอล (HTML) ซึ่งสามารถใช้เทคนิคที่เรียกว่าเฮชทีเอ็มแอลดอมพาเซอร์ (HTML Dom Parser) ในการเข้าไปดึงเอาข้อมูลของเอชทีเอ็มแอลที่ต้องการได้ ในที่นี้ใช้เครื่องมือที่เรียกว่า พีเฮชพี ซิมเปิ้ล เฮชทีเอ็มแอลดอมพาเซอร์ (PHP Simple HTML Dom Parser)

```
<div class="mediaPortrait">
 ▼ <div class="profilePicContainer">
       <div class="blackBackground"> </div>
       <img class="photo
       img" width="75" height="75" title="Fernando 9
       Torres" src="https://scontent.fbkk2-1.fna.fbcdn.net
       /hprofile-xft1/v/t1.0-1/p80x80
       /10577119_527624714005417_4536269760692316328_n.jpg?oh=be16
       068a9130b050fa15fe3040d29ce9&oe=56269772">
       <div class="likeButtonContainer"> </div>
   </div>
 ▶ <a class="mediaRowItem" href="https://www.facebook.com/Torres">
</div>
```

รูปที่ 14 แสดงโค้ดเฮชทีเอ็มแอลของเพจที่ติดตาม

จากรูปที่ 14 เห็นได้ว่ามีข้อมูลที่สำคัญที่สามารถเรียกได้จากเฮชทีเอ็มแอล ให้ดูตรงส่วนของ href ซึ่งเป็นลิงก์ไปยังเพจ ในที่นี้คือเพจที่ชื่อว่า Torres ส่วนท้ายสุดของลิงก์เรียกว่า page_id สำหรับบางเพจที่ตั้งค่าชื่อผู้ใช้แล้ว ส่วนนี้จะเป็น user_name ซึ่งสามารถนำมาใช้แทนกันได้ ให้เก็บข้อมูล page_id เพื่อนำไปใช้ในการเรียกข้อมูลของเพจในภายหลัง

ใช้ เคิล (CURL) ที่เป็นเครื่องมือของพีเฮชพีในการเรียกข้อมูลเพจที่ผู้ใช้งานติดตามทั้ง 400 คน เพื่อนำข้อมูลของเพจมาเป็นข้อมูลตั้งต้นในการหาโพสต์ในขั้นตอนถัดไป

3.3 หา 10 โพสต์ล่าสุดของแต่ละเพจ

ในขั้นตอนนี้จะต้องหาข้อมูลโพสต์ล่าสุดของเพจที่ได้มา เก็บข้อมูล 10 โพสต์ล่าสุดของทุกเพจ เพื่อเตรียมพร้อมสำหรับการหาผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจจากโพสต์ทั้ง 10 โพสต์

วิธีในการหาโพสต์จากเพจนั้นสามารถหาได้จากวิธีในข้อ 3.1 โดยการเรียกไปที่กราฟเอพีไอ https://graph.facebook.com/v2.0/[page_id]?field=posts&access_token=[access-token] โดยเปลี่ยน page_id ไปเรื่อยๆจนครบทุกเพจที่มีอยู่ บางเพจอาจมีข้อมูลของโพสต์ไม่ครบ เช่นเพจที่โพสต์ข้อความไม่ถึง 10 โพสต์ หรือเพจที่โพสต์ข้อความล่าสุดมีอายุนานเกินไป กราฟเอพีไอจะไม่ให้ข้อมูลของโพสต์นั้นออกมา

3.4 หาผู้ใช้งานที่มีปฏิสัมพันธ์กับโพสต์ทั้งหมด

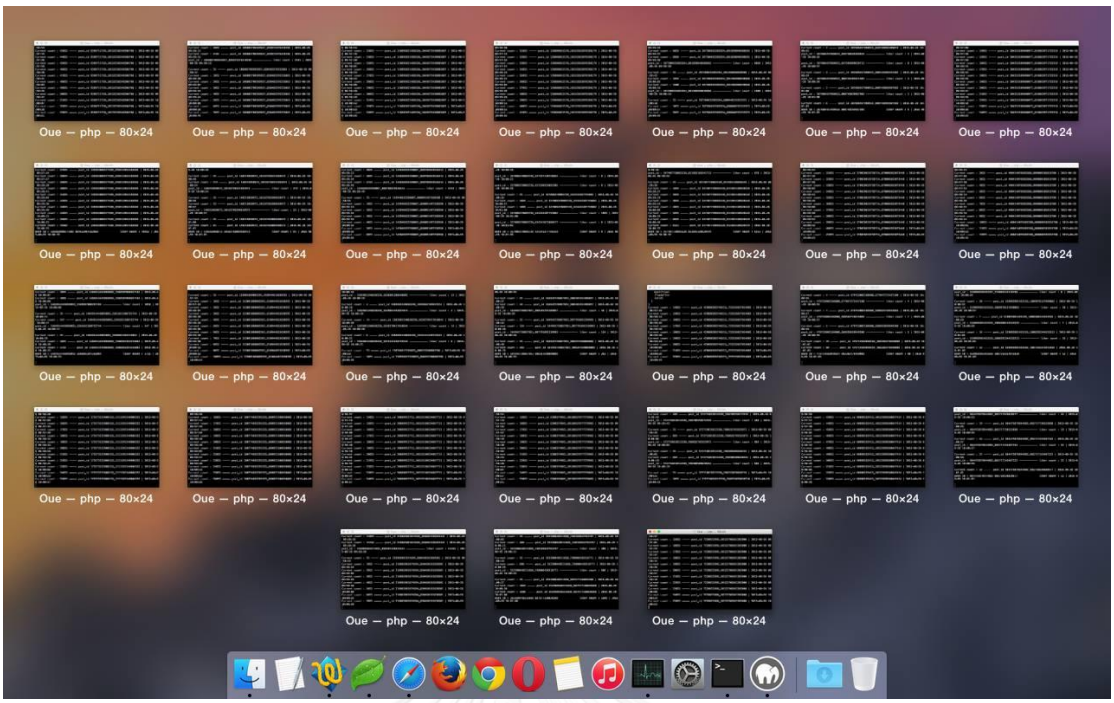
ในขั้นตอนนี้ทำเพื่อหาตัวเชื่อมความสัมพันธ์ระหว่างเพจ เพราะเพจแต่ละเพจไม่สามารถมีความสัมพันธ์ในลักษณะของเพื่อนได้ จึงต้องใช้ผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจเป็นตัวเชื่อมความสัมพันธ์ หากมีผู้ใช้งานที่มีปฏิสัมพันธ์ร่วมกันมาก ย่อมที่จะมีความสัมพันธ์ที่ใกล้กันมาก

การหาผู้ใช้งานที่มีปฏิสัมพันธ์นั้น หาได้จากการดูโพสต์ของเพจว่ามีใครบ้างที่กดไลค์ หรือโต้ตอบข้อความ ในที่นี้ขอเรียกผู้ใช้งานเหล่านั้นว่าผู้กดไลค์ และผู้กดคอมเมนต์ ซึ่งผู้ใช้งาน 2 ประเภทนี้ต้องหามาจากกราฟเอพีไอ https://graph.facebook.com/v2.0/[post_id]?access_token=[access-token] ดังที่เคยอธิบายไว้แล้วในข้อ 3.1 ให้ใส่ post_id จากฐานข้อมูลที่เก็บไว้ วนไปเรื่อยๆ จนได้ข้อมูลของผู้กดไลค์ และ ผู้กดคอมเมนต์จนครบ

ขั้นตอนนี้ค่อนข้างใช้เวลาในการเก็บข้อมูลนานมาก เนื่องจากจำนวนผู้กดไลค์ และผู้กดคอมเมนต์ ที่เรียกมาได้แต่ละครั้งจากกราฟเอพีไอสูงสุดที่ 1,000 คน หากมีโพสต์ไหนเป็นที่นิยม เช่นโพสต์ของ Vin Diesel ดังรูปที่ 15 ที่มีผู้กดไลค์กว่า 7 ล้านคน และผู้กดคอมเมนต์ 1 แสนคน โปรแกรมจะต้องเรียกข้อมูลเป็นจำนวนถึง 7,100 รอบจึงจะเรียกข้อมูลส่วนนี้ได้ 1 โพสต์

รูปที่ 15 แสดงจำนวนผู้กดไลค์ และผู้กดคอมเมนต์ของโพสต์

จากรูปที่ 15 เห็นได้ว่าโพสต์ของดารา บุคคลที่มีชื่อเสียง หรือบริษัทที่มีการจัดกิจกรรมออนไลน์ด้วยเฟซบุ๊ก มีผู้ใช้งานให้ความสนใจเป็นจำนวนมาก การที่จะเก็บข้อมูลเหล่านี้ให้ครบจึงเป็นเรื่องที่ท้าทายมาก ผู้วิจัยจึงเขียนโปรแกรมเพื่อเก็บข้อมูลโดยแยกออกเป็นหลายๆงานให้ทำงานได้พร้อมกัน เขียนด้วยภาษาพีเอชพีแต่สั่งให้ทำงานด้วยคอมมานด์ไลน์ (command line) ของแมคโอเอสเอ็กซ์ (Mac OSX) สั่งให้ทำงานพร้อมกันทั้ง 31 ตัว ซึ่งก่อนที่จะแยกงานทำ ได้มีการระบุถึงส่วนที่ทำงานแต่ละชิ้นในฐานข้อมูล ดังรูปที่ 16



รูปที่ 16 แสดงหน้าจอที่สั่งให้พีเอชพีเก็บข้อมูลพร้อมกัน

ด้วยข้อจำกัดเรื่องประสิทธิภาพของคอมพิวเตอร์ จึงเกิดคอขวดที่มายเอสคิวแอล (MySQL) ที่ใช้เก็บข้อมูล ทำให้เก็บข้อมูลได้ไม่เร็วเท่าที่ควร ขั้นตอนนี้จึงใช้เวลาในการเก็บข้อมูลกว่า 3 สัปดาห์ จึงเก็บข้อมูลได้ครบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.5 คำนวณหาระยะห่างระหว่างเพจทั้งหมด

การค้นหาความสัมพันธ์ของผู้ใช้งานเฟซบุ๊กโดยใช้ความสัมพันธ์แบบเพื่อนในการเชื่อมโยงความสัมพันธ์ เพื่อดูว่าแต่ละคนมีความชอบอะไรที่เหมือนกัน เพื่อทำการแนะนำ หรือนำเสนอสิ่งที่ผู้ใช้งานแต่ละคนน่าจะชื่นชอบได้อย่างถูกต้อง และน่าสนใจ ในมุมมองของเพจเอง ที่ส่วนใหญ่แล้วใช้แทนบุคคล เว็บไซต์ หรือผลิตภัณฑ์ต่างๆ น่าจะถูกมองให้มีรูปแบบความสัมพันธ์ได้เช่นกัน เช่นเพจที่เกี่ยวข้องกับอะไรที่เหมือนกัน ย่อมต้องมีความสัมพันธ์อะไรบางอย่างที่คล้ายกัน ผู้วิจัยจึงนำเอาคนเป็นตัวเชื่อมความสัมพันธ์ระหว่างแฟนเพจเข้าด้วยกัน โดยการลากเส้นเชื่อมโยงระหว่างแฟนเพจโดยใช้คนที่มีปฏิสัมพันธ์กับเพจนั้นเป็นผู้เชื่อมความสัมพันธ์ ก่อเกิดเป็นโครงสร้างเครือข่ายความสัมพันธ์ที่ขยายวงกว้างไปเรื่อยๆขึ้นอยู่กับข้อมูลที่เก็บได้ ดังรูปที่ 17

รูปที่ 17 แสดงความสัมพันธ์ของแฟนเพจโดยใช้คนเป็นตัวเชื่อม

การหาระยะทางระหว่างเพจโดยใช้คนที่มีปฏิสัมพันธ์กับเพจเป็นตัวเชื่อมความสัมพันธ์นั้น เมื่อคนที่กดไลค์เป็นตัวเชื่อมความสัมพันธ์มากขึ้น เท่ากับว่าระยะทางระหว่างเพจคู่นั้นๆจะยิ่งใกล้กันมากขึ้นตามไปด้วย ส่งผลให้เราสามารถแบ่งกลุ่มด้วยวิธีการคลัสเตอร์โดยใช้ระยะทางระหว่างเพจมาคำนวณได้ หากเราสามารถเก็บข้อมูลความสัมพันธ์ของแฟนเพจมากขึ้น จะทำให้ได้กลุ่มของคลัสเตอร์ที่ถูกต้องมากขึ้นเช่นกัน ซึ่งเราสามารถนำข้อมูลเหล่านี้ มาแนะนำเพจที่น่าสนใจที่ยังไม่เคยพบเห็นมาก่อนให้กับผู้ใช้งานในแต่ละกลุ่มได้อีกด้วย

เมื่อสามารถเชื่อมความสัมพันธ์ระหว่างเพจด้วยการใช้คนที่มีปฏิสัมพันธ์ร่วมกันเป็นตัวเชื่อมความสัมพันธ์ได้แล้ว ต้องจัดกลุ่มของเพจเพื่อดูว่าควรจะแนะนำเพจที่อยู่ในกลุ่มใดให้กับผู้ใช้งาน ซึ่งหัวใจของการจัดกลุ่มของเพจได้นั้นคือการหาระยะทางระหว่างเพจให้ได้

ผู้วิจัยมีวิธีการในการหาระยะทางระหว่างเพจต่างๆโดยใช้ความสัมพันธ์ร่วมระหว่างเพจ ซึ่งระยะทางระหว่างเพจ คือ ผลหารระหว่าง จำนวนคนที่มีปฏิสัมพันธ์กับเพจที่น้อยที่สุด กับ จำนวนคนที่มีปฏิสัมพันธ์ร่วมระหว่างเพจ ตามสูตรด้านล่างนี้

$$\text{Distance} = \frac{\text{Min}(P_1,P_2)}{\text{Intersect}(P_1,P_2)}$$

ให้

Distance	=	ระยะทางระหว่างเพจ
P_1	=	เพจที่ 1
P_2	=	เพจที่ 2
$\text{Min}(P_1,P_2)$	=	จำนวนคนที่มีปฏิสัมพันธ์กับเพจที่น้อยที่สุด
$\text{Intersect}(P_1,P_2)$	=	จำนวนคนที่มีปฏิสัมพันธ์ร่วมกันระหว่างเพจ

ขั้นตอนนี้ต้องหาระยะห่างระหว่างเพจทุกคู่เพจ เพื่อง่ายต่อการคำนวณในขั้นตอนถัดไป ซึ่งจริงๆแล้วการแบ่งคลัสเตอร์แบบเค-มีนส์ไม่จำเป็นต้องหาระยะทางทุกคู่ แต่ด้วยการใช้งานเทคโนโลยี

ที่ต่างกัน จึงมีความสะดวกมากกว่าที่หาระยะทางทั้งหมดมาก่อน แล้วจึงส่งข้อมูลทั้งหมดไปในขั้นตอนถัดไป

เนื่องจากจำนวนเพจที่เก็บข้อมูลได้มีจำนวนทั้งสิ้น 80,884 เพจ จะต้องคำนวณทั้งสิ้น $\frac{n * (n-1)}{2}$ = 3.27 พันล้านครั้ง อีกทั้งข้อมูลความสัมพันธ์ที่เกิดขึ้นมากว่า 400 ล้านความสัมพันธ์ ซึ่งเกินกว่าที่เครื่องคอมพิวเตอร์ 1 เครื่องจะสามารถประมวลผลออกมาได้ จึงต้องใช้เครื่องมือที่มีการทำงานในลักษณะของคอมพิวเตอร์หลายเครื่องทำงานพร้อมๆกัน หรือเรียกว่าคลัสเตอร์มาช่วยในการทำงาน

อปาเช สปาร์ค (Apache Spark) เป็นเครื่องมือหนึ่งที่เป็นที่นิยมในการประมวลผลข้อมูลระดับบิ๊กดาต้า (Big Data) ที่มีความเร็วสูง รองรับการทำงานของภาษาไพทอน (Python) และสกาลาร์ (Scalar) ที่สำคัญรองรับการประมวลผลด้วยคลาวด์ประมวลผลแบบยืดหยุ่นของอมาซอน (Amazon Elastic Compute Cloud) หรือเรียกว่า อีซีทู (EC2) ซึ่งการทำงานกับอีซีทูที่เป็นการประมวลผลบนคลาวด์มีความสะดวกเป็นอย่างมาก เนื่องจากไม่ต้องติดตั้งเครื่องคอมพิวเตอร์เอง ประหยัดค่าใช้จ่ายได้เป็นอย่างมาก

การสั่งให้สปาร์คทำงานร่วมกับอีซีทูได้นั้นต้องดาวน์โหลดสปาร์คมาติดตั้งก่อนที่ https://spark.apache.org/downloads.html จากนั้นต้องสมัครเป็นสมาชิกกับอมาซอนเซิร์ฟเวอร์เพื่อใช้งาน EC2

การติดตั้งสปาร์คบนอีซีทูไม่จำเป็นต้องสร้างเครื่องเซิร์ฟเวอร์ด้วยตัวเอง เพียงเรียกใช้คำสั่ง launch ของสปาร์คก็สามารถสร้างเซิร์ฟเวอร์ได้ด้วยตัวเอง อีกตั้งยังติดตั้งเครื่องมือที่จำเป็นบนเครื่องต่างๆให้ด้วย สามารถตั้งค่าได้ว่าจะสร้างกี่เครื่อง ให้ใช้เครื่องที่มีสเปคแบบไหนก็ทำได้ โดยการใช้เทอร์มินอล (Terminal) ของแมคโอเอสเอ็กซ์ (Mac OSX) เข้าไปที่ spark/ec2 จากนั้นพิมพ์คำสั่ง

./spark-ec2 -k <keypair> -i <key-file> -z <zone> --copy-aws-credentials --instance-type=<instance-type> -s <num-slaves> launch <cluster-name>

โดยที่

<keypair>	คือ ชื่อของคีย์แพร์ (keypair) ที่ได้มาจากอีซีทู
<key-file>	คือ ตำแหน่งที่ตั้งของไฟล์คีย์แพร์
<zone>	คือ เครื่องที่จะสร้างนี้อยู่ในโซนไหน เช่น Verginia, Singapore
<instance-type>	คือ เครื่องที่ต้องการใช้งานว่าจะใช้แบบใด สามารถดูรายละเอียดของเครื่องได้จาก http://aws.amazon.com/ec2/instance-types/ หากไม่ใส่ข้อมูลในส่วนนี้ จะได้เป็น instance-type

m1.large ซึ่งมีหน่วยความจำหลัก 7.5 กิกะไบต์ และมี 2 คอร์ ในการประมวลผล

<num-slaves> จำนวนเครื่องที่ใช้ในการประมวลผลข้อมูล

<cluster-name> ชื่อของคลัสเตอร์

งานวิจัยนี้ใช้เครื่องที่เป็น m1.large จำนวน 18 เครื่องในการประมวลผลข้อมูล โดยที่ข้อมูลนำเข้าสำหรับสปาร์คจะเป็นข้อมูลประเภทไฟล์ข้อความ ซึ่งไฟล์ข้อความนี้มีขนาดใหญ่ จึงต้องใช้บริการเก็บรักษาไฟล์อย่างง่ายของอมาซอน (Amazon Simple Storage Service) หรือเรียกว่า เอสทรี (S3) ในการเก็บไฟล์ข้อความ ทำให้ง่ายต่อการเข้าถึงไฟล์จากทุกๆที่

รูปแบบของไฟล์ข้อความที่ใช้ในการหาระยะห่างระหว่างเพจนั้น ได้ออกแบบมาให้อยู่ในรูปแบบที่ 1 แถวของไฟล์ข้อความเป็นตัวแทนของ 1 เพจ ข้อมูลตัวแรกเป็น page_id ข้อมูลตัวถัดไปเป็น user_id ซึ่งถูกคั่นด้วยช่องว่างเพื่อง่ายในการเขียนโปรแกรมแยกข้อความ เช่น

page_id_1 1603690109913686 722049871251547 11755939348321

page_id_2 212495-345757 834948526792738 930989383730242

page_id_3 145111814532103 12840304989901 3948372048371002

เมื่อเตรียมข้อมูลนำเข้าเสร็จเรียบร้อย จึงเขียนโปรแกรมรับข้อมูลไฟล์ข้อความนี้ไปประมวลผล โดยจุดประสงค์คือการหาระยะทางระหว่างเพจ โดยมีข้อมูลส่งออกเป็นไฟล์ข้อความที่คำนวณระยะทางระหว่างเพจเสร็จแล้ว เพื่อนำไปใช้ในกระบวนการถัดไป

ข้อมูลส่งออกมาลักษณะข้อมูลเป็น tuple ที่แต่ละตำแหน่งของข้อมูลมีความหมายที่แตกต่างกัน ตัวอย่างข้อความส่งออกเช่น

(u’page_id_1’, u’page_id_2’, 5,1, 5.0)

(u’page_id_1’, u’page_id_3’, 5,1, 5.0)

(u’page_id_2’, u’page_id’1’, 5, 1, 5.0)

ผู้วิจัยได้กำหนดความหมายของข้อมูลเป็นดังนี้

ข้อมูลตำแหน่งที่ 1 แทน เพจที่ 1

ข้อมูลตำแหน่งที่ 2 แทน เพจที่ 2

ข้อมูลตำแหน่งที่ 3 แทน จำนวนผู้ใช้งานที่มีปฏิสัมพันธ์ที่น้อยกว่าระหว่างเพจที่ 1 กับ 2

ข้อมูลตำแหน่งที่ 4 แทน ปฏิสัมพันธ์ร่วมกันระหว่างเพจ 2 เพจ

ข้อมูลตำแหน่งที่ 5 แทน ระยะห่างระหว่างเพจที่คำนวณได้

การคำนวณระยะทางระหว่างเพจทั้งหมดนี้ ใช้เวลาประมาณ 20 ชั่วโมง ด้วยเครื่องจำนวน 18 เครื่อง ที่มีหน่วยความจำหลัก 7.5 กิกะไบต์ และมี 2 คอร์ ในการประมวลผลข้อมูล

3.6 การแบ่งคลัสเตอร์แบบเค-มีนส์

เมื่อได้ข้อมูลส่งออกจากขั้นตอนที่แล้ว จึงสามารถแบ่งคลัสเตอร์แบบเค-มีนส์จากระยะทางระหว่างเพจที่มีอยู่ทั้งหมดได้ ในขั้นตอนนี้ผู้วิจัยได้นำข้อมูลระยะทางทั้งหมดเข้าฐานข้อมูลที่เป็นมายเอสคิวแอลเพื่อสะดวกต่อการค้นหา และเขียนโปรแกรมด้วยภาษาพีเอชพีในการทำคลัสเตอร์

เริ่มต้นทำด้วยการสุ่มจุดศูนย์กลางของกลุ่ม หรือเรียกว่าเซ็นทรอยด์จำนวน 10 เซ็นทรอยด์ โดยเลือกเพจที่มีปฏิสัมพันธ์ในระดับที่เกินค่าเฉลี่ยขึ้นไป คำนวณจาก

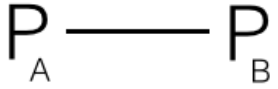
$$\text{ค่าเฉลี่ย} = \frac{\text{จำนวนปฏิสัมพันธ์ทั้งหมด}}{\text{จำนวนเพจทั้งหมด}}$$

$$\text{ค่าเฉลี่ย} = \frac{412{,}741{,}845}{80{,}884}$$

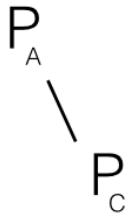
$$\text{ค่าเฉลี่ย} = 5{,}102.89$$

ดังนั้นเราควรหาเพจที่มีจำนวนปฏิสัมพันธ์ที่มากกว่า 5,102 เพราะถ้าเลือกจำนวนปฏิสัมพันธ์น้อยๆ จะได้เพจที่ไม่มีความสัมพันธ์เชื่อมโยงไปยังเพจอื่นๆเลย ทำให้กลุ่มที่ออกมาไม่สามารถกระจายตัวไปหาความสัมพันธ์ได้ดี

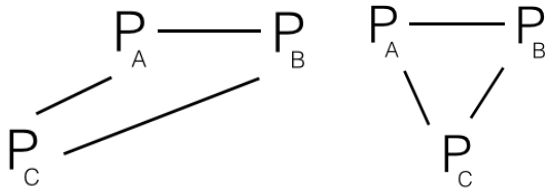
หลังจากนั้นจึงทำการแบ่งคลัสเตอร์แบบเค-มีนส์ ในทางทฤษฎีนั้นการแบ่งคลัสเตอร์แบบเค-มีนส์จะหยุดก็ต่อเมื่อเซ็นทรอยด์ไม่มีการเปลี่ยนแปลง หรือเปลี่ยนแปลงได้น้อยที่สุด แต่ความสัมพันธ์ระหว่างเพจ เป็นความสัมพันธ์ไม่มีทิศทางที่คำนวณได้แน่นอนเช่น เมื่อเพจ A ใกล้กับ B แล้ว ถ้าเพจ C ใกล้กับ A จะไม่สามารถสรุปได้ว่าเพจ B กับ C ใกล้กัน ตามรูปที่ 18-20



รูปที่ 18 แสดงเพจ A และ B ใกล้กัน



รูปที่ 19 แสดงเพจ A และ C ใกล้กัน

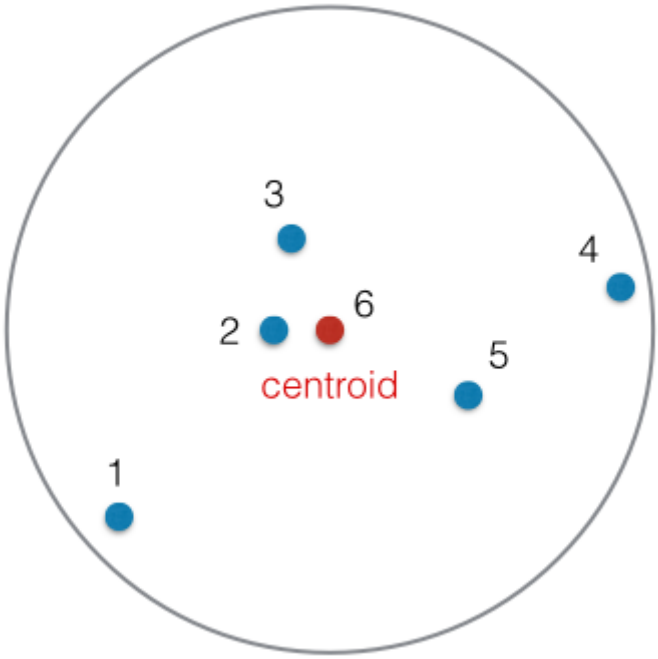


รูปที่ 20 แสดงว่าเพจ B กับ C อาจจะใกล้ หรือไกลกันก็ได้

จากรูปด้านบน ทำให้เราไม่สามารถสรุปตำแหน่งที่แน่ชัดของแต่ละเพจได้ จึงไม่สามารถทำการหยุดคลัสเตอร์ด้วยตำแหน่งได้ จึงกำหนดจำนวนรอบในการทำคลัสเตอร์เพื่อจัดกลุ่มของเพจแทน ในงานวิจัยนี้กำหนดจำนวนการทำงานของคลัสเตอร์ไว้ที่ 50 รอบ

วิธีการหาเซ็นทรอยด์ใหม่ในแต่ละรอบ ทำได้โดยการหาค่าเฉลี่ยระยะทางจากเซ็นทรอยด์กับสมาชิกในคลัสเตอร์ทุกตัว เพื่อคำนวณหาว่า ระยะทางเฉลี่ยระหว่างเซ็นทรอยด์กับสมาชิกทุกตัวคือเท่าใด และสมาชิกตัวใดที่มีระยะทางใกล้กับค่าเฉลี่ยนั้นที่สุด สมาชิกตัวนั้นจะเป็นเซ็นทรอยด์ใหม่

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 21 แสดงตัวอย่างการหา centroid ใหม่

จากรูปที่ 21 พบว่าคลัสเตอร์นี้มีสมาชิกทั้งหมด 6 ตัว โดยจุดที่เป็นสีแดงคือเพจที่ทำหน้าที่เป็นเซ็นทรอยด์ หากวัดระยะเฉลี่ยระหว่างเพจที่ 1 ถึง 5 กับเซ็นทรอยด์มีระยะทางตามข้อมูลด้านล่างนี้

ระยะทางระหว่างเพจที่ 1 กับเซ็นทรอยด์คือ 7

ระยะทางระหว่างเพจที่ 2 กับเซ็นทรอยด์คือ 1

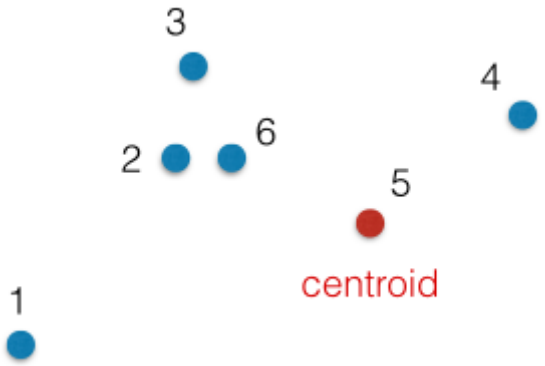
ระยะทางระหว่างเพจที่ 3 กับเซ็นทรอยด์คือ 3

ระยะทางระหว่างเพจที่ 4 กับเซ็นทรอยด์คือ 7

ระยะทางระหว่างเพจที่ 5 กับเซ็นทรอยด์คือ 4

คำนวณค่าเฉลี่ยระยะทาง $= \frac{7+1+3+7+4}{5} = 4.4$

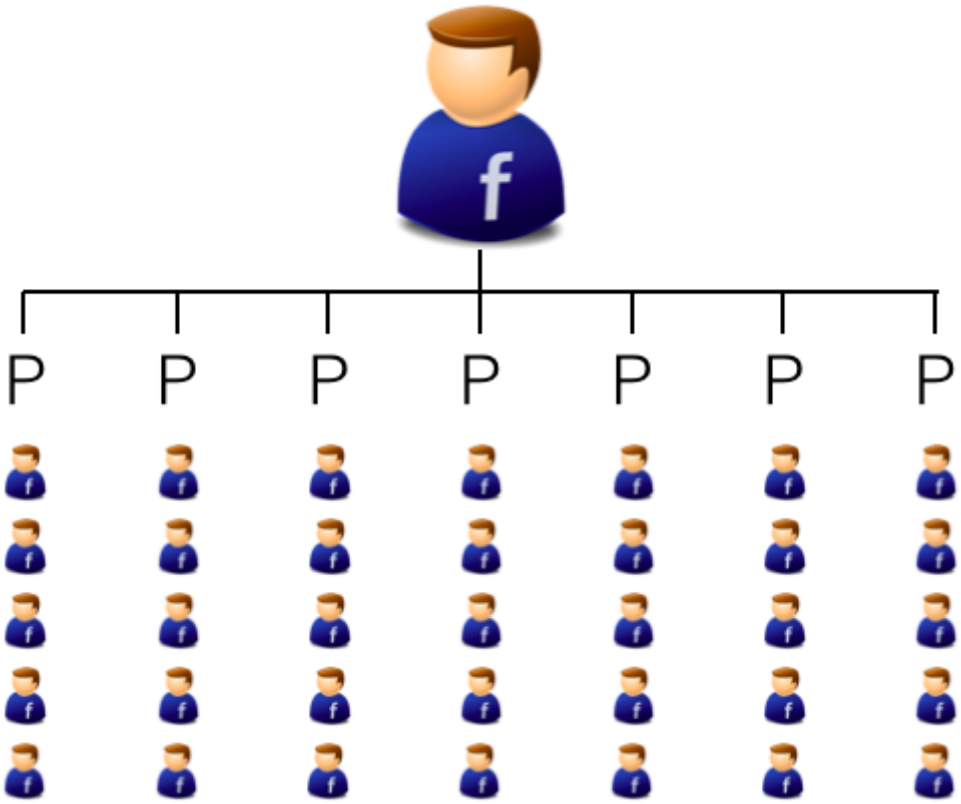
ค่าเฉลี่ยระยะทางระหว่างสมาชิกทุกตัวกับเซ็นทรอยด์คือ 4.4 ซึ่งมีค่าที่ใกล้เคียงที่สุดกับระยะทางระหว่างเพจที่ 5 กับเซ็นทรอยด์นั่นก็คือค่า 4 ดังนั้นเพจที่ 5 จะต้องเป็นเซ็นทรอยด์ใหม่ของคลัสเตอร์นี้ ดังรูปที่ 22



รูปที่ 22 แสดง centroid ใหม่ที่ได้จากคำนวณ

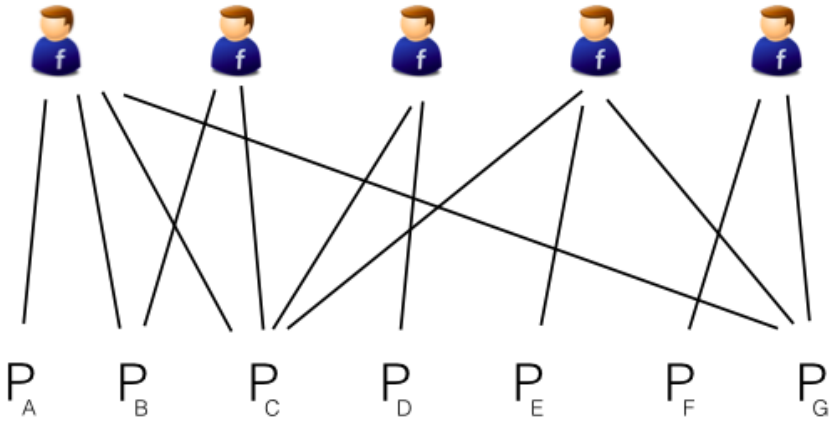
3.7 การใช้ผู้ใช้งานผู้นำในการแนะนำเพจ

วิธีหนึ่งในการแนะนำเพจให้กับผู้ใช้งานนั้น ปกติจะใช้เพจผู้ใช้งานติดตามร่วมกันมากที่สุดมาใช้ในการแนะนำ ผู้ใช้งานที่กล่าวถึง คือผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจจากผู้ใช้งานที่ต้องการค้นหาเพจในการแนะนำ ดังรูปที่ 23



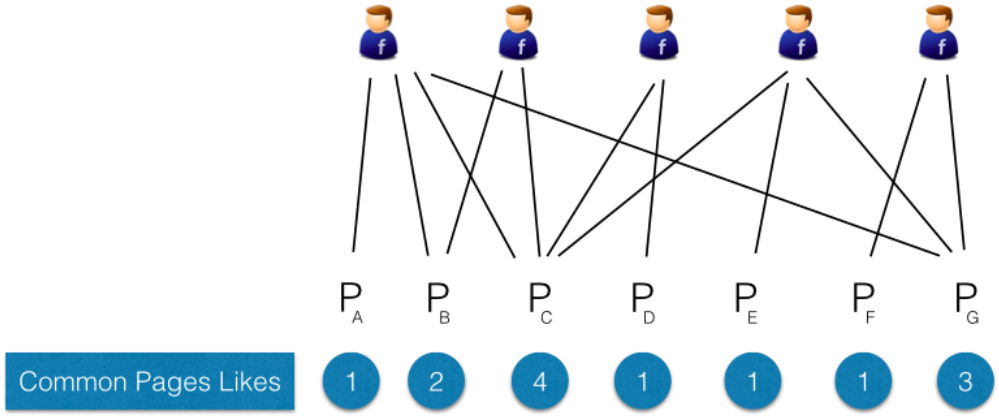
รูปที่ 23 แสดงผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจ

โดยการนำผู้ใช้งานดังกล่าวมาหาเพจที่ผู้ใช้งานติดตาม เพื่อหาเพจที่ติดตามร่วมกัน โดยมีขั้นตอนตามรูปที่ 24



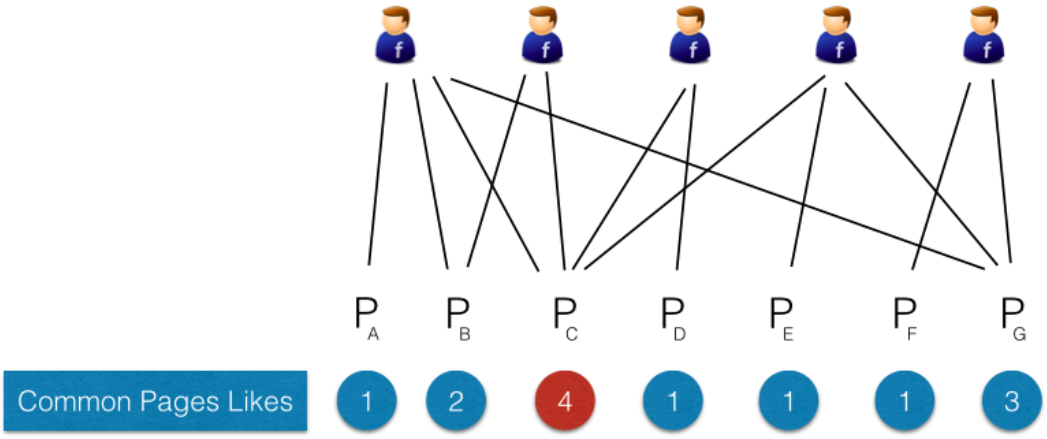
รูปที่ 24 แสดงการติดตามเพจร่วมกันของผู้ใช้งาน

เมื่อภาพของการติดตามเพจร่วมกันเป็นเช่นนี้ จึงต้องนับจำนวนผู้ใช้งานที่ติดตามแต่ละเพจเพื่อใช้ในการหาเพจที่มีผู้ติดตามร่วมกันเป็นจำนวนมากที่สุด ดังรูปที่ 25



รูปที่ 25 แสดงการนับจำนวนผู้ใช้งานที่ติดตามแต่ละเพจ

เมื่อสามารถนับจำนวนผู้ติดตามในแต่ละเพจได้แล้ว ให้แนะนำเพจที่มีจำนวนผู้ใช้งานติดตามร่วมกันมากที่สุดตามรูปที่ 26

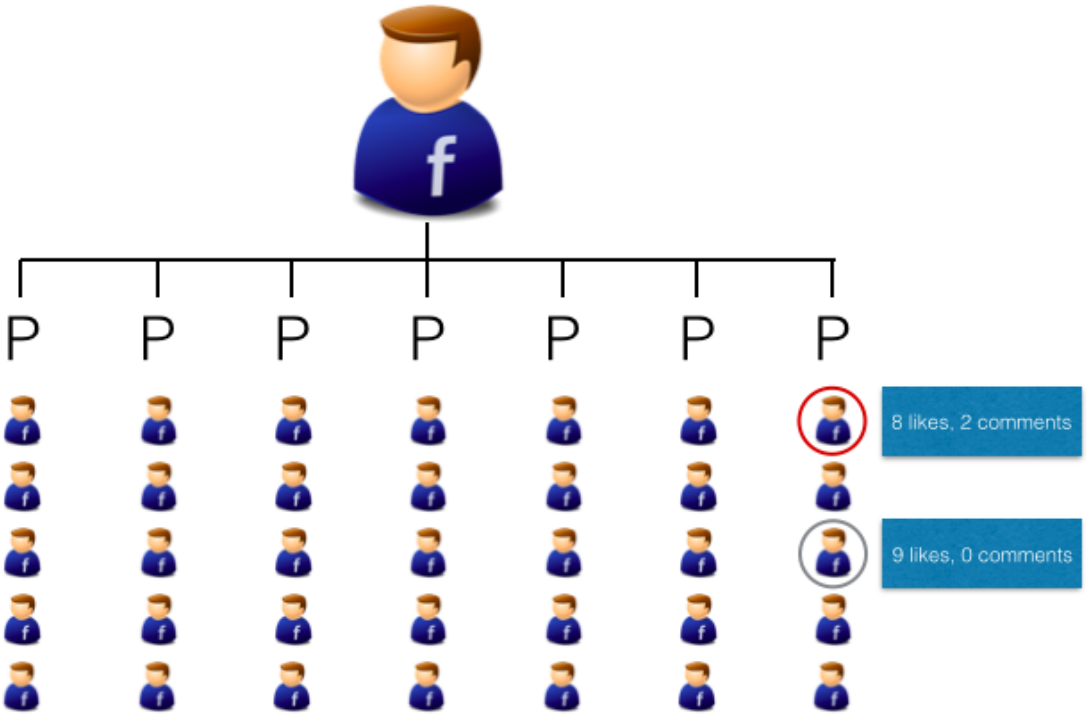


รูปที่ 26 แสดงเพจที่มีผู้ติดตามร่วมกันมากที่สุด

จากรูปที่ 26 จึงแนะนำเพจ C ให้กับผู้ใช้งานที่นำมาทดสอบ เพราะมีจำนวนผู้ติดตามร่วมกันมากที่สุด กระบวนการด้านบนทั้งหมดนี้คือการแนะนำเพจด้วยวิธีพื้นฐาน

ผู้วิจัยได้นำผู้ใช้งานผู้นำเข้ามามีส่วนร่วมในการแนะนำผู้ใช้งาน ซึ่งผู้ใช้งานผู้นำคือผู้ใช้งานที่มีจำนวนปฏิสัมพันธ์มากที่สุดในเพจ เมื่อผู้ใช้งานที่มีความสนใจเพจมาก ย่อมจะมีจำนวนปฏิสัมพันธ์ภาพ และผู้ใช้งานคนนั้นย่อมต้องมีปฏิสัมพันธ์จำนวนมาก หรือชื่นชอบเพจที่มีเนื้อหาแบบเดียวกันกับเพจที่มีปฏิสัมพันธ์มากเช่นกัน ดังนั้นเราจึงใช้ประโยชน์จากความสัมพันธ์ดังกล่าวมาใช้ในการแนะนำเพจ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



รูปที่ 27 แสดงการหาผู้ใช้งานผู้นำ

จากรูปที่ 27 จะเห็นว่าผู้ใช้งานคนหนึ่งมี 8 ไลค์, 2 คอมเมนต์ หมายความว่ามีจำนวนปฏิสัมพันธ์เท่ากับ 10 แต่ผู้ใช้งานอีกคนหนึ่งมี 9 ไลค์, 0 คอมเมนต์ จึงมีจำนวนปฏิสัมพันธ์รวมกันเป็น 9 หากเปรียบเทียบผู้ใช้งานทั้ง 2 คนนี้ สรุปได้ว่า ผู้ใช้งานที่มี 10 ปฏิสัมพันธ์จะเป็นผู้ใช้งานผู้นำของเพจนั้น

เมื่อประยุกต์การใช้งานข้างต้น จึงไม่จำเป็นต้องใช้ผู้ใช้งานทุกคนในการหาเพจที่ติดตามร่วมกัน เพียงใช้แต่ผู้ใช้งานผู้นำก็เพียงพอแล้ว เพราะผู้ใช้งานคนอื่นอาจไม่ได้มีความชื่นชอบในเนื้อหาของเพจมากเท่ากับผู้ใช้งานผู้นำ

3.8 การแนะนำเพจโดยเปรียบเทียบกันระหว่างวิธีต่างๆ

การแนะนำเพจในงานวิจัยนี้ เหลือเพียงใช้ประโยชน์จากคลัสเตอร์ และผู้ใช้งานผู้นำในการแนะนำเพจให้กับผู้ใช้งานเฟซบุ๊ก ในงานวิจัยครั้งนี้ได้ทดลองเปรียบเทียบ 3 วิธีด้วยกัน โดยแต่ละวิธีมีขั้นตอนในส่วนแรกที่เหมือนกันดังนี้

1. การเลือกผู้ใช้งานทดสอบในการหาเพจแนะนำ

จากการเก็บข้อมูลจากเฟซบุ๊กที่เริ่มต้นจากผู้ใช้งาน 400 คน และหาเพจที่ผู้ใช้งานติดตาม จากนั้นจึงหาผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจดังกล่าว ได้ผู้ใช้งานจำนวน 106,638,220 คน ซึ่งการทดลองทั้ง 3 วิธีนี้จะใช้ผู้ใช้งานทดสอบ 400 คนจากจำนวน 106,638,220 คน

2. เลือกเพจทดสอบ และเพจที่ใช้ตรวจสอบ

หาเพจที่ผู้ใช้งานทดสอบติดตาม และแบ่งเพจออกเป็น 2 กลุ่ม คือกลุ่มทดสอบ และกลุ่มที่ใช้ตรวจสอบ แบ่งในอัตราส่วน 80/20 เช่น ผู้ใช้งาน A ติดตามเพจจำนวน 200 เพจ แบ่งเพจเหล่านั้นออกเป็น 2 กลุ่ม คือกลุ่มทดสอบ 80% = 160 เพจ และกลุ่มที่ใช้ตรวจสอบ 20% = 40 เพจ ให้นำเพจกลุ่มทดสอบเข้าสู่ขั้นตอนต่อไป ส่วนกลุ่มที่ใช้ตรวจสอบให้นำไปใช้ตรวจสอบผลลัพธ์ความถูกต้องในขั้นตอนสุดท้าย

3. หาผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจกลุ่มทดสอบ

จากเพจกลุ่มทดสอบจำนวน 80% ให้หาผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจกลุ่มทดสอบทั้งหมด

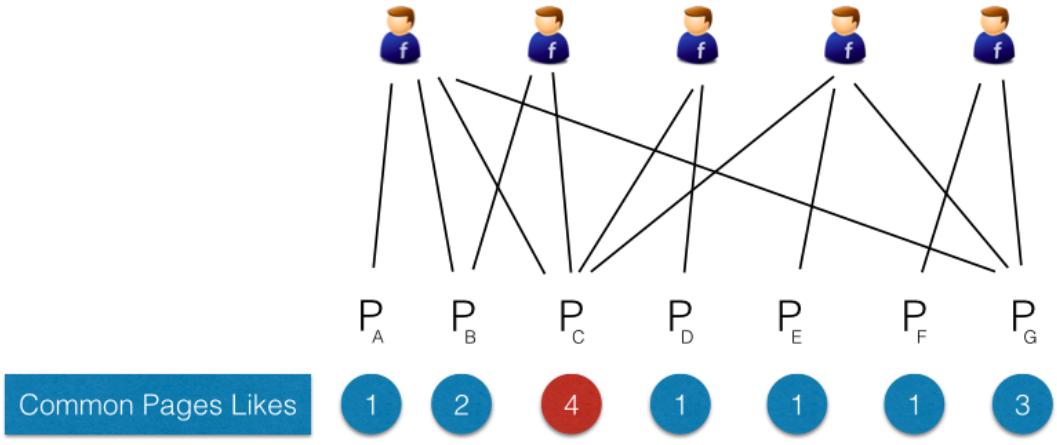
ขั้นตอนการเตรียมข้อมูลข้างต้น ถูกใช้เพื่อนำข้อมูลทั้งหมดไปใช้ในการทดลองทั้ง 3 รูปแบบ ซึ่งรายละเอียดดังนี้

3.8.1 การแนะนำโดยใช้ความชอบร่วมกัน (Common Interest Recommendation)

วิธีนี้เป็นขั้นพื้นฐานของการแนะนำเพจ โดยดูจากความชอบร่วมกันของผู้ใช้งานที่มีต่อเพจของผู้ใช้งานทดสอบ โดยมีขั้นตอนการแนะนำดังนี้

1. ตรวจสอบว่าผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจกลุ่มทดสอบนั้น ติดตามเพจใดบ้าง เพราะเพจที่ผู้ใช้งานติดตาม คือเพจที่มีเนื้อหาที่ผู้ใช้งานสนใจ

2. หาเพจที่ผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจทดสอบนั้นติดตามร่วมกันมากที่สุดมาแนะนำ



รูปที่ 28 แสดงเพจที่ผู้ใช้งานติดตามร่วมกัน

จากรูปที่ 28 พบว่าเพจ C มีการติดตามร่วมกันที่มากที่สุด จึงเลือกเพจ C ในการแนะนำให้กับผู้ใช้งาน

3.8.2 การแนะนำโดยใช้ผู้ใช้งานผู้นำ (Pure Lead Users Recommendation)

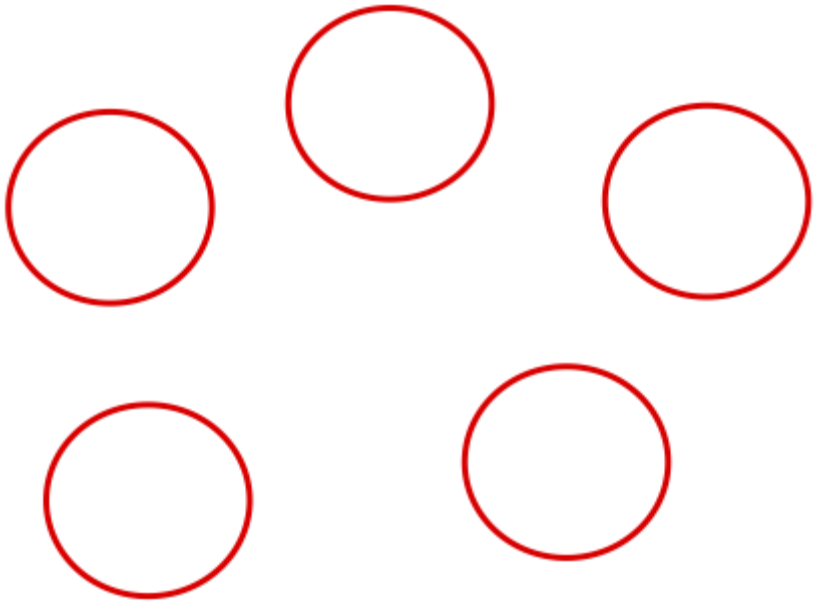
การแนะนำด้วยวิธีนี้เพิ่มผู้ใช้งานผู้นำเข้ามาเพื่อช่วยในการแนะนำ ด้วยสมมติฐานที่ว่า “ถ้าผู้ใช้งานคนใดมีปฏิสัมพันธ์กับเพจใดมากๆ ผู้ใช้งานคนนั้นย่อมต้องมีปฏิสัมพันธ์กับเพจประเภทเดียวกันมากเช่นเดียวกัน” ซึ่งมีขั้นตอนการทดลองดังนี้

1. หาผู้ใช้งานผู้นำจากผู้ใช้งานที่มีปฏิสัมพันธ์มากที่สุด 5 อันดับของเพจกลุ่มทดสอบ
2. ในขั้นตอนนี้เหมือนกับวิธีการแนะนำโดยใช้ความชอบร่วมกันที่หาเพจที่มีการติดตามร่วมกันมากที่สุด ต่างกันตรงที่ไม่ได้ใช้ผู้ใช้งานทุกคนในการหาเพจที่ติดตามร่วมกัน แต่ใช้เพียงผู้ใช้งานผู้นำซึ่งมีจำนวนน้อยกว่าในการหาเพจที่ติดตามร่วมกัน
3. หาเพจที่ผู้ใช้งานผู้นำของเพจทดสอบนั้นติดตามร่วมกันมากที่สุดมาแนะนำ

3.8.3 การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์ (Lead Users and Clusters Recommendation)

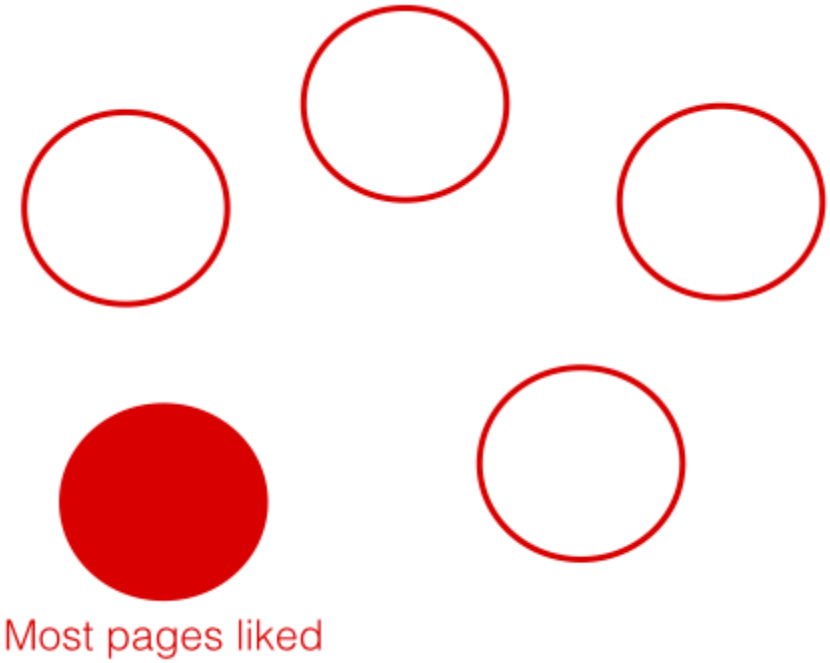
เมื่อได้ผลการแนะนำที่มาจากการจัดกลุ่มของเพจต่างๆแล้ว จะพบว่าอันดับต้นๆของผลการค้นหามักเป็นเพจที่ค่อนข้างทั่วไปที่มีผู้ใช้งานจำนวนมากติดตามอยู่แล้ว ในกรณีของคนไทยเพจที่ปรากฏในส่วนนี้คือ เพจ ICHITAN (https://www.facebook.com/ichitan) เพราะเป็นเพจที่ทำกิจกรรมแจกของรางวัลอยู่บ่อยครั้ง และการทำกิจกรรมของเพจจะเป็นลักษะให้ ไลค์และแชร์เพื่อลุ้นรับของรางวัล จึงทำให้มีผู้สนใจเข้าร่วมกิจกรรมเป็นจำนวนมาก ไม่ได้หมายความว่าเพจไม่ดี หรือไม่มีความสำคัญ แต่ในเรื่องของการแนะนำจะไม่มีความสำคัญเลย เพราะใครๆต่างก็ไลค์เพจนี้อยู่แล้วไม่ว่าจะอยู่กลุ่มใดก็ตาม ซึ่งมีขั้นตอนการทดลองดังนี้

1. หาคลัสเตอร์จากเพจทั้งหมดที่หาได้จากกลุ่มทดสอบในขั้นตอนที่ 3.1 ซึ่งมีจำนวนทั้งหมด 80,884 เพจ หาจำนวน 10 คลัสเตอร์แต่ในภาพที่อธิบายนี้ขออธิบายโดยใช้ภาพที่มีจำนวน 5 คลัสเตอร์ เพื่อให้ง่ายต่อการทำความเข้าใจ ดังรูปที่ 29



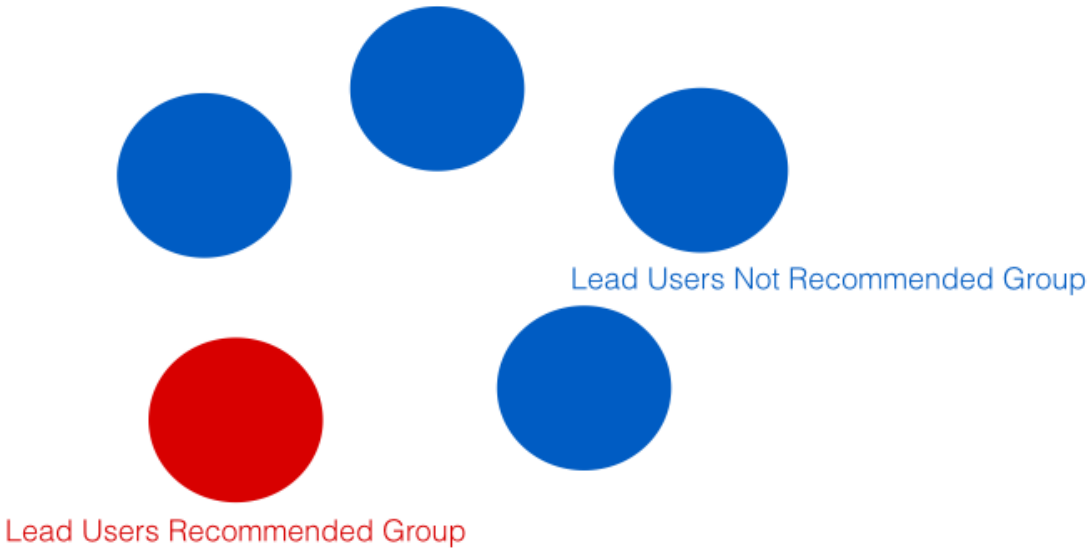
รูปที่ 29 แสดงคลัสเตอร์ที่หามาได้

2. นำเพจจากกลุ่มทดสอบที่มีจำนวน 80% มาตรวจสอบดูว่าเพจดังกล่าวนั้นอยู่ในคลัสเตอร์ใดบ้าง แล้วหาคลัสเตอร์ที่มีจำนวนเพจจาก 80% มากที่สุด ให้แยกเป็นกลุ่มหนึ่ง ส่วนอีกกลุ่มคือคลัสเตอร์อื่นๆ



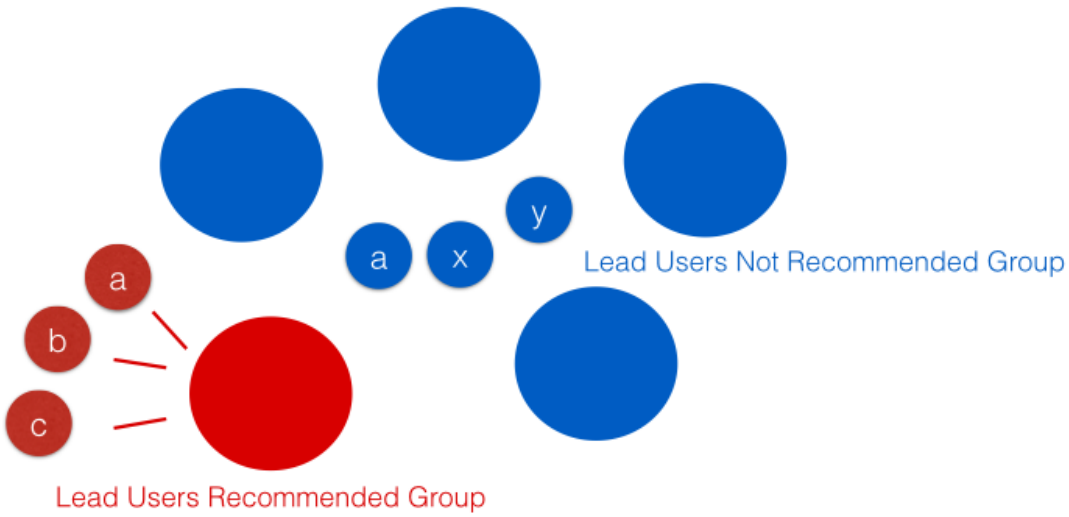
รูปที่ 30 แสดงการแยกคลัสเตอร์ที่หามาได้ออกเป็น 2 กลุ่ม

จากรูปที่ 30 จะเห็นว่าเพจกลุ่มแรกที่มีเพจจาก 80% มากที่สุดเป็นสีแดง ส่วนที่เหลือเป็นกลุ่มที่ไม่มีสี จากนั้นให้ตั้งชื่อกลุ่มใหม่ เป็นกลุ่มที่มีชื่อว่ากลุ่มที่ผู้ใช้งานผู้นำแนะนำ (lead users recommended group) และกลุ่มที่ผู้ใช้งานผู้นำไม่แนะนำ (lead users not recommended group) ดังรูปที่ 31



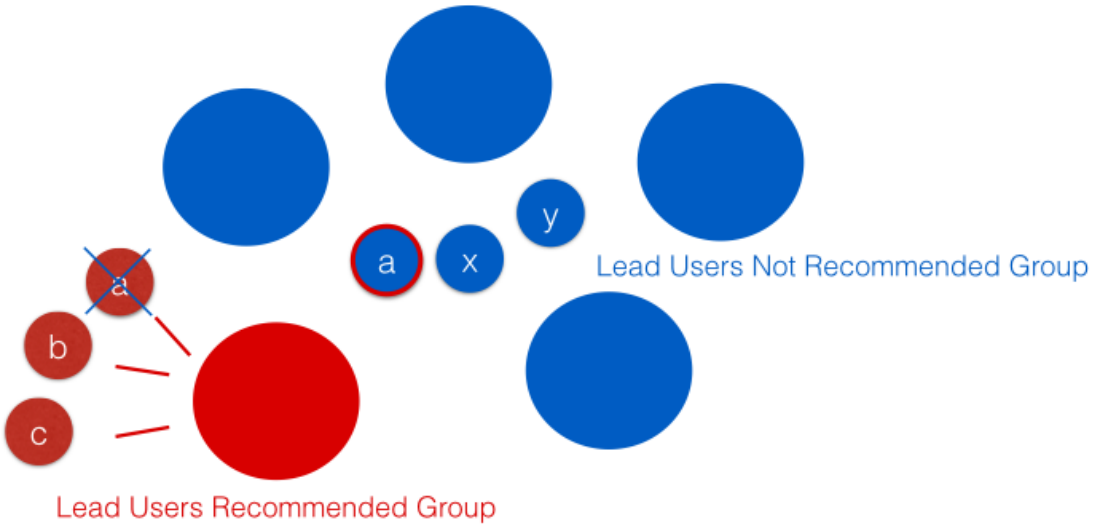
รูปที่ 31 แสดงการตั้งชื่อกลุ่มใหม่

3. ผู้วิจัยใช้ประโยชน์จากผู้ใช้งานผู้นำในการแนะนำเพจ โดยใช้ผู้ใช้งานผู้นำจากเพจทดสอบที่อยู่ในกลุ่มสีแดง มาหาเพจที่ใช้ในการแนะนำ และใช้ผู้ใช้งานผู้นำจากเพจทดสอบที่อยู่ในกลุ่มสีน้ำเงิน มาหาเพจที่ไม่ใช้ในการแนะนำ ดังรูปที่ 32



รูปที่ 32 หาเพจจากผู้ใช้งานผู้นำทั้ง 2 กลุ่ม

4. จากนั้นหาเพจที่ร่วมกันระหว่างสีแดงและสีน้ำเงิน เพื่อนำเพจที่ไม่แนะนำออกไปจากกลุ่มสีแดง เพราะเพจใดก็ตามที่ปรากฏอยู่ในทุกกลุ่ม ย่อมเป็นเพจที่ไม่มีความน่าสนใจในแง่ของการแนะนำ จึงต้องนำออกไปด้วยวิธีนี้



รูปที่ 33 แสดงการตัดเพจที่ไม่แนะนำออกจากผลการแนะนำ

จากรูปที่ 33 เห็นว่าเพจที่ถูกตัดออกไปจากผลการแนะนำคือเพจ a เนื่องจากเพจ a ไปปรากฏอยู่ในกลุ่มสีน้ำเงิน ซึ่งเป็นกลุ่มที่ไม่แนะนำ

3.9 การวัดผล

การวัดผลความถูกต้องของผลลัพธ์การแนะนำสามารถทำได้ด้วยการแบ่งเพจของผู้ใช้งานทดสอบออกเป็น 2 ส่วนคือ เพจที่นำไปค้นหาความสัมพันธ์ และเพจสำหรับวัดผลการแนะนำ เนื่องจากผลการแนะนำที่ถูกต้อง ย่อมตรงกับความชอบของผู้ใช้งาน จึงใช้เพจที่ผู้ใช้งานติดตามอยู่แล้วมาเพื่อวัดผลการแนะนำ ผู้จัดทำแบ่งออกเป็นสัดส่วนดังนี้

1. เพจที่นำไปค้นหาความสัมพันธ์จำนวน 80%

2. เพจสำหรับวัดผลการแนะนำจำนวน 20%

ผลการแนะนำที่ได้มาจะนำมาเปรียบเทียบกับเพจจำนวน 20% เพื่อวัดอัตราความถูกต้อง ในที่นี้ใช้วัดจำนวนฮิทเรต (hit rate) ว่าใน 20 อันดับสูงสุดที่แนะนำ มีเพจจำนวนกี่เปอร์เซนต์ที่ตรงกับความชอบของผู้ใช้งาน

$$\text{Hit rate} = \frac{UN}{RN} * 100$$

ให้

Hit rate	=	เปอร์เซนต์ความถูกต้องของผลลัพธ์การแนะนำ
UN	=	จำนวนการแนะนำเพจที่ถูกต้องเปรียบเทียบกับเพจจำนวน 20% ของผู้ใช้งานที่นำมาวัดผล
RN	=	ในที่นี้คือ 20 เพจ เป็นเพจ 20 อันดับแรกที่นำมาเปรียบเทียบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
ผลการทดลอง

บทนี้กล่าวถึงผลการทดลองของงานวิจัยชิ้นนี้ และกล่าวถึงข้อมูลต่างๆที่เก็บเอามาวิเคราะห์เพื่อนำมาทำเป็นผลการทดลองนี้ ซึ่งประกอบไปด้วย

4.1 ข้อมูลเพจที่ผู้ใช้งานตัวอย่างติดตาม

งานวิจัยชิ้นนี้เริ่มต้นโดยเก็บข้อมูลการกดติดตามเพจของผู้ใช้งานเฟซบุ๊กจำนวน 400 คน ซึ่งในทางสถิติแล้ว ตัวอย่างผู้ใช้งานเฟซบุ๊กจำนวน 400 คนถือว่าเพียงพอแล้วที่จะเป็นตัวแทนขอผู้ใช้งานเฟซบุ๊กจำนวน 1.4 พันล้านคน ด้วยความผิดพลาดที่ยอมรับได้ที่ 5%

การเก็บข้อมูลเพจที่ผู้ใช้งานทั้ง 400 คนกดติดตามจากการเก็บข้อมูลโดยการใช้เฮชทีเอ็มแอลดอมพาเซอร์เก็บข้อมูลจากหน้าโปรไฟล์ของผู้ใช้งานได้เพจรวมกันทั้งหมด 80,884 เพจ จำแนกตามผลลัพธ์ตามตารางที่ 3 และ 4

ตารางที่ 3 แสดงรายละเอียดของเพจที่ผู้ใช้งานตัวอย่าง 400 คนกดติดตาม

รายละเอียด	**จำนวนเพจ**
เพจที่ไม่มีผู้ใช้งานกดติดตามซ้ำกัน	62,021 (76.68%)
เพจที่มีผู้ใช้งานกดติดตามซ้ำกัน	18,863 (23.32%)
เฉลี่ยจำนวนเพจที่ผู้ใช้งานกดติดตามต่อ 1 คน	202.21

ตารางที่ 4 แสดงเพจที่ผู้ใช้งานตัวอย่างติดตามมากที่สุด

อันดับ	**ชื่อเพจ**	**ผู้ใช้งานตัวอย่างที่ติดตาม**	**เปอร์เซนต์**	**ผู้ใช้งานที่ติดตามทั้งหมด**
1	ตัน ภาสกรนที	269	67.25	11,452,612
2	SCB Thailand	187	46.75	2,901,357
3	ICHITAN	175	43.75	3,577,352
4	พระมหาวุฒิชัย วิชิรเมธี	163	40.75	5,096,055
5	KBank Live	156	39.00	1,868,021
6	Samsung Mobile Thailand	146	36.50	3,911,559

7	พระมหาสมชาย ฐานวุฑโฒ	144	36.00	5,090,979
8	KFC	141	35.25	730,934
9	Lazada	136	34.00	12,013,355
10	Oishi Drink Station	126	31.50	3,305,052
11	บิณฑ์ บรรลือฤทธิ์	125	31.25	5,735,606
12	Major Cineplex Group (Thailand)	123	30.75	2,317,893
13	Samsung Thailand	121	30.25	1,849,783
14	GTH	117	29.25	2,897,028
15	Welovetogo	117	29.25	1,909,156
16	May Feungarom	112	28.00	3,704,641
17	Khaosod	110	27.50	5,624,333
18	Tesco Lotus – เทสโก้ โลตัน	105	26.25	1,486,973
19	KUTE CLUB	103	25.75	2,418,006
20	WORKPOINT ENTERTAINMENT	103	25.75	3,605,831

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 ข้อมูลโพสต์

จากข้อมูลผู้ใช้งานตัวอย่างเริ่มต้นที่ 400 คน ได้จำนวนเพจที่กดติดตามจำนวน 80,884 เพจ ผู้วิจัยทำการเก็บโพสต์ 10 โพสต์ล่าสุด เพื่อนำมาวิเคราะห์ข้อมูล พบว่ามีจำนวนโพสต์ 643,157 โพสต์ แสดงว่ามีบางเพจที่ไม่มีโพสต์ หรือมีโพสต์ไม่ถึง 10 โพสต์ หรือเป็นไปได้ว่าเป็นโพสที่เก่าเกินไป ทำให้กราฟเอฟีไอไม่สามารถดึงข้อมูลออกมาแสดงได้

ตารางที่ 5 แสดงจำนวนปฏิสัมพันธ์ที่มีต่อเพจ เรียงจากมากไปน้อย

เพจ	จำนวนปฏิสัมพันธ์
Vin Diesel	2,076,512
Neha Sharma	2,067,721
Madhuri Dixit – Nene	2,049,439
UEFA Champions League	1,495,360

Piyanut Pannoy (ปิยะนุช แป้นน้อย)	1,470,531
Karim Benzema	1,396,160
Deepika Padukone	1,385,478
Lker Casillas	1,300,853
Sri Sri Radha Rasabihari Temple	1,294,590
Huge Jackman	1,278,183
Paul Walker	1,274,861
Khatu Shyam ji (Shyam Baba)	1,253,876
Boo	1,249,888
พุทธสามัคคี	1,189,516
Avril Lavigne	1,170,101
Justin Bieber	1,161,426
Humans of New York	1,145,173
Zac Efron	1,104,608
Gareth Bale	1,066,470
Paul Walker	1,019,792

จากตารางที่ 5 พบว่า ถึงแม้เพจที่ผู้ใช้งานตัวอย่างติดตามเป็นอันดับต้นๆ แต่ก็ไม่พบเลยว่าอันดับ 1 ของเพจที่มีผู้ใช้งานตัวอย่างติดตาม จะมีปฏิสัมพันธ์เป็นอันดับต้นๆในตารางที่ 3 อาจเป็นเพราะว่าเพจที่มีผู้ใช้งานติดตามมาก ไม่ได้มีเนื้อหาที่ผู้ใช้งานสนใจติดตามก็เป็นไปได้ หรืออาจเพราะมีการเล่นกิจกรรมเพื่อแจกของรางวัล จึงมีผู้ใช้งานติดตามเพียงเพื่อรับทราบข่าวสาร

ตารางที่ 6 แสดงจำนวนปฏิสัมพันธ์ที่มีต่อโพสต์

ไลค์	คอมเมนต์
401,202,093	11,539,752

จากตารางที่ 6 พบว่าการมีปฏิสัมพันธ์ของผู้ใช้งานกับโพสนั้น ไลค์มีอัตราส่วนมากกว่าคอมเมนต์ถึงร้อยละ 3,376.70 ซึ่งนับว่าเป็นจำนวนที่แตกต่างกันมาก

4.3 ข้อมูลความสัมพันธ์ที่เกิดขึ้น

การเก็บข้อมูลมาทั้งหมดจากผู้ใช้งานตัวอย่าง 400 คน ใช้เวลา 1 เดือน เก็บข้อมูลตั้งแต่วันที่ 1 มิถุนายน พ.ศ.2558 ถึงวันที่ 30 มิถุนายน พ.ศ.2558 ได้ข้อมูลโดยรวมดังนี้

ตารางที่ 7 แสดงข้อมูลโดยรวมจากการเก็บข้อมูลทั้งหมด

ข้อมูล	จำนวน
จำนวนเพจทั้งหมด	80,884 เพจ
จำนวนโพสต์ทั้งหมด	643,157 โพสต์
จำนวนผู้ใช้งานที่มีปฏิสัมพันธ์ทั้งหมด	106,638,220 คน
จำนวนปฏิสัมพันธ์ของผู้ใช้งานกับเพจ	411,509,369 ปฏิสัมพันธ์

จากการทำคลัสเตอร์เริ่มต้นด้วยการใช้ 10 เซ็นทรอยด์ และทำทั้งหมด 50 รอบ สามารถแจกแจงผลการทำคลัสเตอร์ออกเป็นดังนี้คือ 50 คู่เพจที่ใกล้กันมากที่สุด และตัวอย่างสมาชิกเพจใน 10 คลัสเตอร์ ดังตารางที่ 8 และ 9

ตารางที่ 8 แสดงตัวอย่างเพจที่ใกล้กันที่สุด 50 คู่แรก

ลำดับที่	เพจ A	เพจ B
1	Tlux51	Bag's house Shop
2	ทรัพย์เจริญ แทรเวล	หจก.รุ่งทวีการท่องเที่ยว
3	ขอบสนาม	ไบรอัน เดอะซวย
4	น้าค่อม ชวนชื่น (fanclub)	แมนยูคลับ
5	We Love Coach	The Peninsula Bangkok
6	Bei Otto	Voucher_travel
7	New page Mookies Zapp	คลิปดังเฟสบุ๊ค
8	ภูฝางทัวร์	เอิร์ธ เบาะซิ่ง
9	บอย ท่าพระจันทร์ FC	Kirk Schiller
10	Sexyเบาเบา	รีทวิต
11	Me Cute By พัชชี่	พาเที่ยวแม่แจ่ม
12	Kipling familyshop	บ้านผึ้งป่านศรนารายณ์

13	DJ Soda Fans	เฟตเฟ่
14	Jo's BAG	Kipling Chicshop
15	HOT Ohozaa	Chodchapak Pholthanachod
16	Iloveyourbag	JAPAN BRAND NAME
17	Cloninggems High End	dngems จิวเวลรี่
18	YouLike V.2	คอบอลNews
19	LV 101	วิถีชนคนเมืองล้านนา
20	Anime-I.Com	เรื่องในห้องแชท
21	ที่นอนจารุภัณฑ์ นุ่นใหม่บริสุทธิ 100% โพธาราม	จักรยานไฟฟ้า HONDA E BIKE
22	สาวก บาร์ซ่า	Fernny Fanclub
23	I AM United กูไม่ได้บ้าแต่กูคือSpartaแห่งUnited	SCG เมืองทอง ยูไนเต็ด
24	Jewelry (จิวลี่) มือสอง by Jib	ชัยนาทโพสต์
25	PunPun Sutatta Udomsilp	เพลินดีนี่ รีสอร์ท นครนายก fanpage
26	เข้าครัว	Kid Fashionista: เสื้อผ้าเด็กน้อย พร้อมส่ง
27	กูว่าแล้วมันต้องยิง	ตัน พาไปสกลที v.2
28	พริ้งกี้แฟชั่นหลากหลาย	จำหน่ายมือถือราคาถูก
29	รอบรั้วรามฯ	พรรคตะวันใหม่ มหาวิทยาลัยรามคำแหง
30	วิจารณ์ มวยไทย 7 สี	ชีวิตยังคงสวยงาม - bodyslam
31	Emotional Moment in Japan	โรงเรียนสารสาสน์เอกตราแผนกอนุบาล
32	หลวงปู่ทองดี อนิโฆ	Watphut Radio FM100.25
33	Pumi allergy ไรฝุ่น ภูมิแพ้ ภูม	Kidshop By Leo
34	คริปเด็ด คริปดัง คริปเสียง เชิญทางนี้	เปิดม่านความจริง [The Truth]

35	แหวนเพชร ทองคำ เครื่องประดับ สินค้าหลุดจำนำ	OK Diamonds
36	BROMPTON THAILAND FAN CLUB	โรงเรียนวิศวกรรมรถไฟ
37	แดงแหนมเนือง	สำนักงาน ออฟฟิต เคลื่อนที่ ร้านกาแฟ สปา
38	คลิปเด็ด กีฬาดัง	Phi Hormones
39	Lovelyshoe99	Laosball
40	หลุยส์ ฟาน กัล	Dragon-King
41	Hinoko shop	หจก.เดอะสตาร์เซรามิค แอนด์ ปริ้นท์
42	กูว่าแล้ว มันต้องเกรียน	Sora Aoi
43	Scuba Soccer & badminton	YUTHA FEDFE FC.
44	พรรคเขียว จรรโลงโลก	ohchang Fasai Resort
45	MERCEDES BENZ OWNER CLUB FRANCE	Brandname Shop
46	รวมพลคนต้านโรคเรื้อนเปียก	Jardin De Mangue Pakchong
47	กระทรวงติ่งเกาหลีแห่งประเทศไทย	ร้านสัก ส เสือ tattoo shop
48	Pure Life Newgen	ผ้าห่มนาโน ผ้าขนหนู ลายลิขสิทธิ์
49	Chockdee jewelry	ใจไหม การ์เม้นท์
50	ปืนสวัสดิการข้าราชการ (สน.สก.)	Palm Lodge Resort - Thailand

ตารางที่ 9 แสดงตัวอย่างสมาชิกเพจในคลัสเตอร์

cluster	เพจ
1	บันทึกหน้าสุดท้าย
	ทุกๆ คนที่สู้ชีวิต
	คาราเต้
	บลูทาวน์ สงขลา
	วัดสุนันทวนาราม

2	30 กำลังแจ๋ว
	รักสามเศร้า
	ก๋วยเตี๋ยวขอสาว
	The Craziest Videos
	อลิส
3	ครีมบำรุงผิว V2 White Perfect
	Mr.Friendly [นาฬิกาแฟชั่น]
	FAR METIC ขาว-ใส ไม่ต้องรองพื้น by JAM
	Beauty-Club
	สมาคมโดนใจ
4	Hotel Liberty
	Eagles Hotel California
	มีดีฮอลิเดย์
	Redseed Pub
	Restaurant Zollhaus
5	Aircraft maintenace engineering
	เมนี คอมพิวเตอร์
	อะโดบี อาฟเตอร์เอฟเฟกต์
	แมคบุ๊กโปร
	กรมส่งเสริมการปกครองท้องถิ่น
6	77 Coffee
	สิ่งเล็กๆที่เรียกว่ารัก
	94 Coffee
	Drinking Beer
	Butterfly club
7	T-Bow – ทีโบว์หมูกะทะ
	Share-Games
	Green wave
	Jiggaban
	Stamp collecting

8	เที่ยวทะเลใต้ ตรัง กระบี่ สตูล ภูเก็ต
	Happy TG Travel
	Travel Tham Sai – ท่องเที่ยวถ้ำสาย
	ท่องเที่ยวคลับ
	ภาษาญี่ปุ่นและสิ่งดีๆ
9	Hentai-tv
	แฮปปี้คนเลี้ยงหมู
	Ragnarok Server Game Online
	ขาวใสสุขภาพดี ใน 7 วัน
	นันเดด
10	เรารักอาสาฯ ราชมงคล
	รับสอนพิเศษภาษาจีนชลบุรี
	Bangkok Christian College
	Saint Mary's School
	โรงเรียนเตรียมอุดมศึกษา

4.4 ผลการทดลอง

การทดลองแบ่งออกเป็น 3 รูปแบบคือ การแนะนำโดยใช้ความชอบร่วมกัน, การแนะนำโดยใช้ผู้ใช้งานผู้นำ และการแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์

4.4.1 การประเมินความถูกต้องของการแนะนำด้วยฮิตเรต

ในการทดลองนี้ใช้ผู้ใช้งานทดสอบจำนวน 400 คนจากผู้ใช้งานที่มีอยู่ในระบบ 106,638,220 คน ทดลองโดยใช้เพจที่ผู้ใช้งานทดสอบแต่ละคนติดตามติดตามจำนวน 80% มาประมวลผลด้วยวิธีการทดลองจากที่กล่าวมา และนำผลของการแนะนำ 20 อันดับแรกมาเปรียบเทียบกับเพจอีก 20% ของผู้ใช้งานทดสอบเพื่อหาฮิตเรตได้ผลตามตารางที่ 10

ตารางที่ 10 แสดงฮิตเรตของการแนะนำเพจด้วยวิธีการทดลองทั้ง 3 รูปแบบ

วิธีการทดลอง	ฮิตเรต
การแนะนำโดยใช้ความชอบร่วมกัน	3.21%
การแนะนำโดยใช้ผู้ใช้งานผู้นำ	15.73%
การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์	16.62%

ทดลองเปลี่ยนข้อมูลในการทดสอบจากเพจ

เพจทดลอง 80% เพจที่ใช้ประเมินผล 20% เป็น

เพจทดลอง 20% เพจที่ใช้ประเมินผล 80%

เพื่อดูว่าการเปลี่ยนสัดส่วนของจำนวนเพจทดสอบกับประเมินผล มีผลต่อฮิตเรตหรือไม่

ตารางที่ 11 แสดงฮิตเรตของการแนะนำเพจด้วยวิธีการทดลองทั้ง 3 รูปแบบ โดยเปลี่ยนสัดส่วนของเพจทดสอบ

วิธีการทดลอง	ฮิตเรต
การแนะนำโดยใช้ความชอบร่วมกัน	4.83%
การแนะนำโดยใช้ผู้ใช้งานผู้นำ	24.16%
การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์	27.20%

จากตารางที่ 10 และ 11 ที่พบว่าการทดลองแบบใช้ผู้ใช้งานผู้นำสามารถเพิ่มประสิทธิภาพในการแนะนำได้เป็นอย่างมาก แต่การใช้คลัสเตอร์นั้น ไม่ได้มีผลมากนักในการเพิ่มประสิทธิภาพในการแนะนำ และการเปลี่ยนสัดส่วนให้เพจที่ใช้ในการทดลองให้ลดลงจาก 80% เป็น 20% ส่งผลให้ ฮิตเรตมีอัตราที่สูงขึ้น

การใช้ผู้ใช้งานผู้นำมาช่วยในการแนะนำเพจนั้นให้ผลลัพธ์ที่ดีมาก แต่ผลลัพธ์ในด้านของความแม่นยำของการทดลองแบบใช้คลัสเตอร์ และไม่ใช้คลัสเตอร์ ไม่มีความแตกต่างกันมาก ผู้วิจัยจึงเข้าไปประเมินผลในส่วนของคำตอบว่าการแนะนำเพจของแต่ละรูปแบบการทดลองนั้นมีความแตกต่างกันอย่างไรบ้าง

ตารางที่ 12 แสดงผลลัพธ์ 10 อันดับแรก เปรียบเทียบระหว่างการใช้กับไม่ใช้คลัสเตอร์ของผู้ใช้งาน 1 คน

การแนะนำโดยใช้ผู้ใช้งานผู้นำ		การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์	
เพจ	จำนวนผู้ติดตาม	เพจ	จำนวนผู้ติดตาม
ตัน ภาสกรนที	11,452,612	Eat All Day	150,830
SCB Thailand	2,901,357	เรียนเจ้านายที่เคารพ	283,248
ICHITAN	3,577,352	เรารักด่านตรวจ	343,648
พระมหาวุฒิชัย วิชิรเมธี	5,096,055	แต่เพื่อเงินฉันทำได้	94,583
KBank Live	1,868,021	IKEA Thailand	331,316
คนอะไรเป็นแฟนหมี	1,393,642	SiamphoneFan	627,270
Lazada	12,013,355	หมอๆ ตะลุยโลก	66,084
May Feungarom	3,704,641	แมนยูคลับ	306,306
Khaosod	5,624,333	English.Today (เรียนภาษาอังกฤษวันละคำ)	310,662
KUTE CLUB	2,418,006	รวมดาวสาว Office	87,271

จากตารางที่ 12 เปรียบเทียบผลการทดลองที่ใช้ผู้ใช้งานผู้นำมาช่วยในการหาคำตอบ พบว่าเพจที่แนะนำบ่อยที่สุดของการทดลองการแนะนำโดยใช้ผู้ใช้งานผู้นำ เป็นเพจที่ค่อนข้างใกล้เคียงกันกับเพจที่มีผู้ใช้ตัวอย่างติดตามมากที่สุด ซึ่งเป็นเพจที่ค่อนข้างจะทั่วไป ผู้ใช้งานหลายๆกลุ่มติดตาม แต่ในฝั่งของเพจแนะนำด้วยวิธีการแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์ ที่ใช้คลัสเตอร์มาช่วยในการหาคำตอบ พบว่าเพจที่แนะนำจะเป็นเพจที่มีคนติดตามระดับกลางๆ มีลักษณะค่อนไปในเรื่องใดเรื่องหนึ่งโดยเฉพาะ จะไม่มีผลลัพธ์ทั่วๆไปที่คนกดติดตามเป็นจำนวนมากออกมา เป็นเพราะว่าคลัสเตอร์ได้กรองข้อมูลส่วนนั้นออกไปแล้ว

บทที่ 5
บทสรุป

งานวิจัยนี้ได้นำเสนอวิธีในการแนะนำเพจของเฟซบุ๊กให้ตรงกับความสนใจของผู้ใช้งาน โดยใช้โครงสร้างความสัมพันธ์ของเพจที่มีปฏิสัมพันธ์ระหว่างผู้ใช้งานกับเพจเป็นตัวเชื่อมความสัมพันธ์เข้าด้วยกัน ยิ่งมีผู้ใช้งานที่มีปฏิสัมพันธ์ร่วมกันระหว่างเพจมากเท่าใด เพจก็ย่อมมีความสัมพันธ์ที่ใกล้กันมากยิ่งขึ้น มีการใช้ผู้ใช้งานผู้นำเพื่อเป็นคนที่จะชี้นำไปยังเพจน่าจะตรงกับความต้องการกับผู้ใช้งานในการแนะนำ และใช้คลัสเตอร์ในการจัดกลุ่มเพจที่มีความใกล้เคียงกันเข้าด้วยกัน เพื่อใช้ประโยชน์จากกลุ่มนี้ในการคัดแยกเพจทั่วไปออกจากผลการแนะนำ

การวิจัยนี้ได้เก็บข้อมูลจากเฟซบุ๊กโดยใช้เวลา 1 เดือน ระหว่างวันที่ 1-30 มิถุนายน พ.ศ. 2558 โดยเก็บข้อมูลเริ่มต้นจากผู้ใช้งานจำนวน 400 คน ได้เพจที่ผู้ใช้งานกดไลค์ 80,884 เพจ ได้โพสต์จากเพจ 643,157 โพสต์ เมื่อเก็บข้อมูลผู้ใช้งานที่มีปฏิสัมพันธ์กับเพจทั้งหมด พบว่ามีจำนวนผู้ใช้งานที่มีปฏิสัมพันธ์ 106,638,220 คน และจำนวนปฏิสัมพันธ์ระหว่างผู้ใช้งานกับเพจคือ 411,509,369 ปฏิสัมพันธ์

การทำวิจัยนี้เปรียบเทียบผลจาก 3 การทดลองด้วยกันคือ การแนะนำโดยใช้ความชอบร่วมกัน, การแนะนำโดยใช้ผู้ใช้งานผู้นำ, การแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์ เพื่อหาว่าผู้ใช้งานผู้นำ และคลัสเตอร์มีผลต่อความถูกต้องในการแนะนำอย่างไร

จากผลการทดลองพบว่าผู้ใช้งานผู้นำมีผลทำให้ความแม่นยำในการแนะนำสูงขึ้น เพราะผู้ใช้งานผู้นำคือผู้ใช้งานที่ค่อนข้างมีปฏิสัมพันธ์กับเพจมาก ย่อมต้องมีความสนใจในเรื่องนั้นมากๆ จึงสามารถแนะนำเพจที่เกี่ยวข้องกันได้ดี ในขณะที่คลัสเตอร์ไม่ทำให้ความแม่นยำมีความแตกต่างกันมากนัก แต่การมีคลัสเตอร์ส่งผลให้ผลลัพธ์ของเพจที่แนะนำมีความเฉพาะเจาะจงในด้านใดด้านหนึ่งมากขึ้น ดังนั้นการแนะนำเพจของเฟซบุ๊กด้วยวิธีการแนะนำโดยใช้ผู้ใช้งานผู้นำและคลัสเตอร์จึงเป็นวิธีที่ได้ผลในการแนะนำดีที่สุด

รายการอ้างอิง

[1] H.-Y. Ho and H.-Y. Pan, "Use behaviors and website experiences of facebook community," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, 2010, pp. V1-379-V1-383.

[2] W. Hsin-Ying, L. Kuan-Liang, and C. Trappey, "Understanding customers using Facebook Pages: Data mining users feedback using text analysis," in *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on*, 2014, pp. 346-350.

[3] R. Nararatwong, N. Pongsupankij, K. Atchariyachanvanich, and N. Cooharojananone, "Factors affecting users' attitude toward Facebook application in Thailand," in *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, 2013, pp. 1054-1059.

[4] D. Terrana, A. Augello, and G. Pilato, "Facebook Users Relationships Analysis Based on Sentiment Classification," in *Semantic Computing (ICSC), 2014 IEEE International Conference on*, 2014, pp. 290-296.

[5] W. Suraworachet, S. Premsiri, and N. Cooharojananone, "The Study on the Effect of Facebook's Social Network Features toward Intention to Buy on F-commerce in Thailand," in *Applications and the Internet (SAINT), 2012 IEEE/IPSJ 12th International Symposium on*, 2012, pp. 245-250.

[6] W. Longfei and L. Nianlong, "Social streams recommendation in sina microblog with relation of user and interest," in *Information Science and Technology (ICIST), 2014 4th IEEE International Conference on*, 2014, pp. 480-483.

[7] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," presented at the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 2010.

[8] T. Otsuka, T. Yoshimura, and T. Ito, "Evaluation of the Reputation Network Using Realistic Distance between Facebook Data," in *Web Intelligence and*

Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on, 2012, pp. 383-388.

[9] Y. Yulong, "A personalized collaborative recommendation algorithm based on user clustering smoothing," in *BioMedical Information Engineering, 2009. FBIE 2009. International Conference on Future*, 2009, pp. 383-385.

[10] T. Yamane, *Statistics; an introductory analysis.* New York: Harper and Row, 1967.

ประวัติผู้เขียนวิทยานิพนธ์

นายเอกสิทธิ์ เพิ่มพูนพิพัฒน์ เกิดวันอาทิตย์ที่ 21 ตุลาคม พ.ศ.2527 สำเร็จการศึกษาระดับปริญญาวิทยาศาสตร์บัณฑิต คณะวิทยาศาสตร์และเทคโนโลยี ภาควิชาวิทยาการคอมพิวเตอร์มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2550 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2556

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University