อภิธันคาวาร

สำนักหอสมุด
สัญญาเลขที่ R2559C228

รายงานวิจัยฉบับสมบูรณ์

โครงการ การแก้ปัญหาข้อมูลไม่เพียงพอต่อการวิเคราะห์ความสนใจของนักท่องเที่ยวโดยใช้ข้อมูลจากเฟสบุ๊ค

คณะผู้วิจัย
ดร.อนงค์พร ไศลวรากุล คณะวิทยาศาสตร์
รองศาสตราจารย์ ดร.ไกรศักดิ์ เกษร คณะวิทยาศาสตร์

สำนักหอสมุด มหาวิทยาลัยนเรศวร
วันลงทะเบียน 31 ก.ค. 2562
เลขทะเบียน 1020159
เลขเรียกหนังสือ QA
76
.9 .I58
อ157ร
2560

สนับสนุนโดยกองทุนวิจัยมหาวิทยาลัยนเรศวร

สารบัญ

สารบัญ ..... ก
สารบัญภาพ ..... ข
บทคัดย่อ ..... ค
บทสรุปสำหรับผู้บริหาร ..... ง
บทที่ 1 บทนำ 1
1.1 ความเป็นมาของปัญหา 1
1.2 วัตถุประสงค์ของการวิจัย 2
1.3 ขอบเขตของงานวิจัย 2
1.4 สมมติฐานของการวิจัย 2
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 3
2.1 องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล 3
2.2 การค้นหาความสนใจของผู้ใช้ (USER INTEREST DETECTION) 4
2.3 การสร้างแบบจำลองความสนใจของผู้ใช้ (USER INTEREST MODELING) 8
2.4 เทคนิคการแนะนำข้อมูลให้ผู้ใช้ (RECOMMENDATION TECHNIQUES) 11
บทที่ 3 วิธีดำเนินการวิจัย 13
3.1 เครื่องมือที่ใช้ในการวิจัย 13
3.2 การเก็บรวบรวมข้อมูล 14
3.3 วิธีการดำเนินงานวิจัย 14
3.4 กรอบแนวคิดการวิจัย 15
3.5 การแนะนำข้อมูลการท่องเที่ยว (RECOMMENDATION) 24
บทที่ 4 ผลการวิจัย 29
4.1 ผลการวิเคราะห์ความต้องการของผู้ใช้ 29
4.2 ผลการออกแบบและพัฒนาระบบ 32
4.3 ผลการประเมินระบบ 36
บทที่ 5 บทสรุป 47
5.1 สรุปผลการวิจัย 47
5.2 ข้อเสนอแนะ 47
บรรณานุกรม 48

สารบัญภาพ

ภาพ 1 องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล 4
ภาพ 2 ตัวอย่างแบบจำลองฐานข้อมูลแบบ Flat File 9
ภาพ 3 ตัวอย่างแบบจำลองฐานข้อมูลสัมพันธ์ 9
ภาพ 4 วิธีการวิเคราะห์ข้อมูลจากเครือข่ายสังคม 16
ภาพ 5 ตัวอย่างฟิลด์ข้อมูลของผู้ใช้ที่สามารถดึงมาจากเฟสบุ๊ค 16
ภาพ 6 การแนะนำข้อมูลการท่องเที่ยว 24
ภาพ 7 แหล่งค้นหาข้อมูลสถานที่ท่องเที่ยวต่าง ๆ 30
ภาพ 8 สัดส่วนการใช้สังคมออนไลน์เพื่อหาข้อมูลการท่องเที่ยว 31
ภาพ 9 ความสำคัญในการพัฒนาระบบการให้ข้อมูลการท่องเที่ยวเฉพาะบุคคล 31
ภาพ 10 การคลิกเพื่อเลือกจุดหมายปลายทางและช่วงเวลา 32
ภาพ 11 หน้าจอแนะนำสถานที่ท่องเที่ยว 32
ภาพ 12 หน้าจอการปรับค่าความสนใจ 33
ภาพ 13 เลือกสถานที่เพื่อสร้างแผนการเดินทาง 34
ภาพ 14 ปุ่มเริ่มประมวลผลข้อมูลจากเฟสบุ๊ค 35
ภาพ 15 ระบบแนะนำสถานที่ท่องเที่ยวไปยังผู้ใช้ 35
ภาพ 16 แบบสอบถามลำดับความสนใจ 39
ภาพ 17 ค่าเฉลี่ยความถูกต้องของลำดับหมวดหมู่ความสนใจ 40
ภาพ 18 ค่าเฉลี่ยความถูกต้องของลำดับหมวดหมู่ความสนใจเมื่อระยะเวลาเซ็คอินต่างกัน 41
ภาพ 19 ค่าเฉลี่ยเวลาที่ใช้ในการประมวลผลเมื่อระยะเวลาเซ็คอินต่างกัน 42
ภาพ 20 Precision@10 และ Rankscore10 ของการแนะนำเปรียบเทียบระหว่างการใช้เกณฑ์ที่ต่างกัน 46

บทคัดย่อ

ปัญหาโคลด์สตาร์ท (Cold start problem) คือปัญหาที่รู้จักกันดีในระบบแนะนำข้อมูล (Recommendation system) ซึ่งหมายถึง ข้อมูลมีไม่เพียงพอต่อการวิเคราะห์ความสนใจของผู้ใช้ในเวลาเริ่มต้นของการใช้งาน ซึ่งปกติมักจะเกิดกับผู้ใช้ใหม่ โดยปกติระบบจะให้ผู้ใช้รายใหม่กรอกข้อมูลเริ่มต้น เช่น ข้อมูลส่วนตัว เพื่อให้ระบบสามารถเริ่มทำการแนะนำสินค้าหรือบริการได้ อย่างไรก็ตาม วิธีการดังกล่าวทำให้เกิดความยุ่งยากต่อผู้ใช้ หรือทำให้ผู้ใช้มีภาระเพิ่มมากขึ้น และส่วนมากผู้ใช้จะไม่ยอมกรอกข้อมูลดังกล่าวเนื่องจากต้องการความปลอดภัยส่วนตน ดังนั้นในงานวิจัยนี้จึงนำเสนอเทคนิคใหม่ในการใช้ประโยชน์จากข้อมูลของเครือข่ายสังคมออนไลน์ เพื่อแก้ปัญหาดังกล่าวที่ต้องให้ผู้ใช้กรอกข้อมูล โดยกรอบแนวคิดงานวิจัยที่นำเสนอนี้จะวิเคราะห์ความสนใจของผู้ใช้จากข้อมูลในเฟสบุ๊ค และหาความสนใจของผู้ใช้เพื่อแนะนำสถานที่ท่องเที่ยวให้กับผู้ใช้ จากการทดลองพบว่า การใช้ข้อมูลเช็คอินของผู้ใช้ สามารถใช้แก้ปัญหาโคลด์สตาร์ทได้อย่างมีประสิทธิภาพ

Abstract

Cold start problem is a well-known problem for recommendation system (RS) when the initial data is not adequate for user preference analysis which usually happen with a new user, so called cold start user problem. Typically, the system has to ask users to input some initial data e.g. personal data to allow the RS to start suggestion some products to a user. However, this method causes additional task to a user and is usually ignored because of security reason. This research presents a novel technique to exploit social network data to solve such a problem. The sophisticated framework is developed to extract user data available from Facebook and analyze user preference in the domain of Tourism and then recommend some interesting attraction to a user. The experimental results illustrates that Check-in data from Facebook is very useful for cold start problem solving and then effectively recommendation performance can be achieved.

บทสรุปสำหรับผู้บริหาร

เครือข่ายสังคมออนไลน์มีบทบาทสำคัญต่อชีวิตประจำวันของนักท่องเที่ยวมากขึ้น และมีอิทธิพลต่อการตัดสินใจของนักท่องเที่ยวเป็นอย่างมาก เครือข่ายสังคมออนไลน์มีการเติบโตมากขึ้นทุกๆ ปี และปัจจุบันมีผู้ใช้นับล้านๆ คนเพื่อค้นหาข้อมูลต่างๆ ที่ตนเองสนใจ และแชร์ข้อมูลดังกล่าวไปยังเพื่อนๆ ของตน เครือข่ายสังคมออนไลน์ที่ได้รับความนิยม ได้แก่ Facebook Foursquare Twitter เป็นต้น ข้อมูลต่างๆ ที่อยู่บนเครือข่ายสังคมออนไลน์ถือเป็นข้อมูลที่มีค่า และสามารถนำไปใช้ประโยชน์ต่อธุรกิจต่างๆ ได้เป็นอย่างดี เช่น การโฆษณาสินค้า หรือประชาสัมพันธ์กิจกรรมต่างๆ

สำหรับการแนะนำข้อมูลให้กับนักท่องเที่ยวนั้นมีปัญหาหนึ่งคือ การมีข้อมูลไม่เพียงพอต่อการวิเคราะห์ความสนใจของนักท่องเที่ยว ซึ่งปัญหาดังกล่าวนี้ได้รับความสนใจจากนักวิจัยทางศาสตร์ของวิทยาการคอมพิวเตอร์เป็นอย่างมาก ดังนั้นงานวิจัยนี้จึงมุ่งเน้นในการแก้ไขปัญหาดังกล่าว โดยการนำข้อมูลจากเครือข่ายสังคมออนไลน์มาใช้ให้เกิดประโยชน์ ในบรรดาข้อมูลต่างๆ ในเครือข่ายสังคมออนไลน์ ข้อมูลการเช็คอิน ถือว่ามีประโยชน์มากต่อธุรกิจการท่องเที่ยว เนื่องจากทำให้ทราบว่า นักท่องเที่ยวมีพฤติกรรมการเดินทางและการท่องเที่ยวอย่างไร แต่ข้อมูลเช็คอินนี้ยังไม่ได้ถูกนำมาวิเคราะห์เพื่อแนะนำสถานที่ท่องเที่ยวอย่างมีประสิทธิภาพมากนัก ดังนั้นในงานวิจัยนี้จึงมีการนำข้อมูลการเช็คอินมาผ่านกระบวนการต่างๆ เพื่อหาความสนใจของผู้ใช้ และแนะนำสถานที่ท่องเที่ยวที่อาจจะตรงกับความสนใจของผู้ใช้ได้

จากการทดลองพบว่า 1) ข้อมูลจากเครือข่ายสังคมออนไลน์มีประโยชน์ในการนำมาวิเคราะห์หาความสนใจของผู้ใช้หรือนักท่องเที่ยวแต่ละคนได้อย่างมีประสิทธิภาพ 2) สามารถลดภาระการสอบถามข้อมูลจากนักท่องเที่ยวได้ ทำให้เกิดความพึงพอใจของผู้ใช้ต่อระบบแนะนำข้อมูลในระดับดีมาก จากการค้นพบดังกล่าวถือเป็นต้นแบบให้กับระบบแนะนำสินค้าอื่นๆ นำแนวคิดไปประยุกต์ใช้ เพื่อแนะนำสินค้าต่างๆ ที่ตรงกับความสนใจของผู้บริโภคได้อย่างมีประสิทธิภาพต่อไปได้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ในระบบแนะนำข้อมูลเฉพาะบุคคล (PRS) นั้นการค้นหาความสนใจของผู้ใช้เป็นหัวใจสำคัญในการที่จะแนะนำผลิตภัณฑ์หรือบริการให้ได้ตรงกับความต้องการของผู้ใช้อย่างแท้จริงได้ อย่างไรก็ตามการค้นหาความสนใจมีความท้าทายอย่างยิ่งสำหรับผู้พัฒนาระบบแนะนำข้อมูลเฉพาะบุคคลคือปัญหาโคลด์สตาร์ท (Cold Start Problem) กล่าวคือการมีข้อมูลไม่เพียงพอในการสรุปความสนใจสำหรับผู้ใช้ที่เริ่มต้นใช้งานระบบ ด้วยความสำคัญของปัญหาดังกล่าวจึงมีหลายแนวคิดที่จะพัฒนาเทคนิคการค้นหาความสนใจที่สามารถแก้ปัญหาข้างต้นและได้ข้อสรุปความสนใจของผู้ใช้ได้อย่างถูกต้องและมีประสิทธิภาพ

เครือข่ายสังคม (Social Network) นอกจากจะช่วยโฆษณาประชาสัมพันธ์การท่องเที่ยวได้แล้วยังเป็นที่เก็บข้อมูลจำนวนมากจากผู้ใช้ประกอบกับการเจริญเติบโตของเครือข่ายสังคมที่เป็นไปอย่างรวดเร็วทำให้ข้อมูลทั้งที่เกิดจากการสนทนา แสดงความคิดเห็น แบ่งปัน มีจำนวนที่มากขึ้น ข้อมูลจำนวนมหาศาลเหล่านี้ยังไม่ถูกนำไปใช้ประโยชน์เท่าที่ควร อย่างไรก็ตามนักวิจัยส่วนหนึ่งพยายามจะใช้ประโยชน์จากข้อมูลเหล่านี้โดยการนำข้อมูลจากเครือข่ายสังคมมาวิเคราะห์เพื่อหาความสนใจของผู้ใช้ เช่น การใช้แท็ก (Tag) ในโซเชียลบุ๊กมาร์ก (Social Bookmarking) หาความสนใจของผู้ใช้เพื่อสร้างคำแนะนำในระบบค้นหาเพลง (Firan, et al., 2007) การใช้ข้อความทวีต (Tweet) ในทวิตเตอร์ (Twitter.com) ค้นหาความสนใจของผู้ใช้เพื่อสร้างระบบแนะนำข่าว (Abel, et al., 2011) เป็นต้น เครือข่ายสังคมจึงเป็นแหล่งข้อมูลที่สามารถนำไปสู่ข้อสรุปความสนใจและสามารถนำมาประยุกต์ใช้ในการพัฒนาระบบแนะนำข้อมูลการท่องเที่ยวแบบเฉพาะบุคคลอย่างมีประสิทธิภาพได้ นอกจากนี้ยังมีแนวโน้มที่จะสามารถแก้ปัญหาโคลด์สตาร์ทและลดภาระของผู้ใช้ที่ต้องทำการกรอกข้อมูลความสนใจด้วยตนเองในระบบเดิมอีกด้วย

ด้วยความสำคัญของปัญหาดังกล่าว ผู้วิจัยจึงได้มีแนวคิดในการพัฒนาระบบแนะนำข้อมูลการท่องเที่ยว เพื่อช่วยนำเสนอข้อมูลสถานที่ท่องเที่ยวให้ได้ตรงกับความสนใจของผู้ใช้อย่างแท้จริง รวมถึงแก้ไขปัญหาสำคัญ คือปัญหาโคลด์สตาร์ท (Cold start) โดยการใช้ประโยชน์จากข้อมูลในเครือข่ายสังคม

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและพัฒนาวิธีการแก้ปัญหาโคลด์สตาร์ทโดยใช้ข้อมูลจากเครือข่ายสังคม
2. เพื่อวิเคราะห์ความสนใจของนักเที่ยวโดยใช้ข้อมูลจากเครือข่ายสังคม

1.3 ขอบเขตของงานวิจัย

ในงานวิจัยนี้จะใช้ข้อมูลการเช็คอินจากเฟสบุ๊ค (Facebook) เพื่อตรวจสอบความสนใจของผู้ใช้ โดยจำนวนผู้ใช้ที่เข้าร่วมทดสอบระบบจะมีจำนวนไม่ต่ำกว่า 100 คน โดยมีความหลากหลายทางด้านอายุ เพศ และการศึกษา โดยผู้ร่วมโครงการทั้งหมดจะมาจากนิสิต และบุคคลากรในมหาวิทยาลัยนเรศวร

1.4 สมมติฐานของการวิจัย

การค้นหาความสนใจเฉพาะบุคคลเป็นส่วนสำคัญในการพัฒนาระบบแนะนำข้อมูลเฉพาะบุคคลโดยข้อมูลที่สกัดจากเครือข่ายสังคมที่มีอยู่จำนวนมากบนอินเทอร์เน็ตอาจใช้เป็นตัวแทนเพื่อแสดงความสนใจของแต่ละผู้ใช้ได้จึงนำไปสู่สมมติฐานดังนี้

สมมติฐานงานวิจัย: ข้อมูลจากเครือข่ายสังคมสามารถนำมาแก้ไขปัญหาโคลด์สตาร์ทได้

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

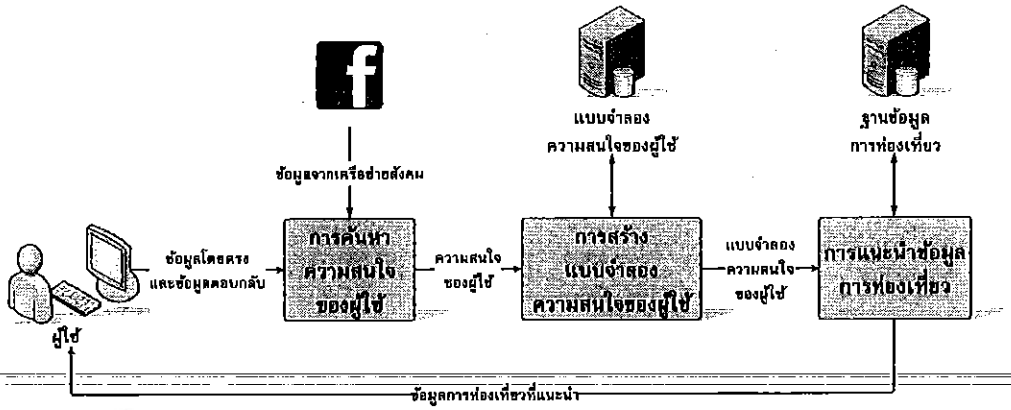
การพัฒนาระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคลนี้ ผู้วิจัยได้ทำการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง ดังนี้

1. องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล
2. เครือข่ายสังคมกับการนำมาใช้ในระบบแนะนำข้อมูลเฉพาะบุคคล

2.1 องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล

ระบบแนะนำข้อมูลเฉพาะบุคคล เป็นระบบที่ใช้ความสนใจของแต่ละผู้ใช้เป็นเงื่อนไขหนึ่งในการพิจารณาข้อมูลเพื่อนำมาแนะนำให้กับผู้ใช้ เช่นเดียวกันกับระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคลเป็นระบบที่นำเสนอข้อมูลการท่องเที่ยวในลักษณะต่าง ๆ เช่น ข้อมูลสถานที่ท่องเที่ยว แผนการเดินทาง ข้อมูลแพ็คเกจท่องเที่ยว เป็นต้น ให้กับนักท่องเที่ยวตามความสนใจที่แตกต่างกันไปในแต่ละบุคคล ลักษณะองค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคลที่แต่ละผู้วิจัยได้พัฒนาขึ้นจะแตกต่างกันไปในรายละเอียด แต่โดยองค์ประกอบหลักแล้วจะประกอบไปด้วย 2 ส่วนที่สำคัญ คือ 1. การค้นหาความสนใจของผู้ใช้ และ 2. การแนะนำข้อมูลการท่องเที่ยวตามความสนใจของผู้ใช้

ระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคลมีองค์ประกอบหลักดังภาพที่ 1 ประกอบด้วย 1) การค้นหาความสนใจของผู้ใช้ (User Interest Detection) เป็นการหาข้อสรุปความสนใจและปรับปรุงให้ทันต่อความสนใจที่เปลี่ยนไปของผู้ใช้ผ่านแหล่งที่มาหลายแหล่ง เช่น ข้อมูลที่ได้จากผู้ใช้โดยตรง ข้อมูลจากเครือข่ายสังคม ข้อมูลตอบกลับ เป็นต้น 2) การสร้างแบบจำลองความสนใจของผู้ใช้ (User Interest Modeling) เป็นการการจัดเก็บข้อมูลความสนใจ ลงในฐานข้อมูล และ 3) การแนะนำสถานที่ท่องเที่ยว (Tourism Attractions Recommendation) เป็นการแนะนำสถานที่ท่องเที่ยวโดยใช้ข้อมูลความสนใจที่ได้มาจากแบบจำลองความสนใจของผู้ใช้



ภาพ 1 องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล

ทั้งนี้ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับองค์ประกอบหลักทั้งสาม โดยจะอธิบายเพิ่มเติมในหัวข้อดังต่อไปนี้

2.2 การค้นหาความสนใจของผู้ใช้ (User Interest Detection)

การค้นหาความสนใจของผู้ใช้เป็นกระบวนการที่สำคัญของระบบแนะนำข้อมูลแบบเฉพาะบุคคลเป็นอย่างยิ่ง เนื่องจากในการที่จะแนะนำผลิตภัณฑ์หรือบริการให้ได้ตรงกับความต้องการของผู้ใช้จะต้องทราบถึงความสนใจที่มีของผู้ใช้ นอกจากนี้ยังต้องเรียนรู้ความสนใจที่เปลี่ยนไปของผู้ใช้ที่เกิดขึ้นได้ตลอดเวลา เพื่อให้ผู้ใช้ได้รับข้อมูลการแนะนำที่มีประสิทธิภาพมากยิ่งขึ้น โดยในกระบวนการนี้ประกอบไปด้วย 2 ส่วน คือ 2.2.1 การได้มาซึ่งข้อมูลความสนใจของผู้ใช้ (User Interest Acquisition) และ 2.2.2 การเรียนรู้ความสนใจของผู้ใช้ (User Interest Learning)

2.2.1 การได้มาซึ่งข้อมูลความสนใจของผู้ใช้ (User Interest Acquisition)

การได้มาซึ่งข้อมูลความสนใจของผู้ใช้ แบ่งเป็น 2 ลักษณะ คือ 1) การได้มาซึ่งข้อมูลความสนใจของผู้ใช้โดยตรง (Explicit) หมายถึง การที่ระบบได้รับข้อมูลจากผู้ใช้โดยตรง เช่น ข้อมูลรายละเอียดส่วนบุคคลที่ผู้ใช้กรอกให้กับระบบเมื่อลงทะเบียน การตอบคำถามของผู้ใช้ การทำแบบสอบถาม เป็นต้น สำหรับ 2) การได้มาซึ่งข้อมูลความสนใจของผู้ใช้โดยอ้อม (Implicit) คือ การที่ระบบทำการเก็บข้อมูลจากการใช้งานหรือพฤติกรรมของผู้ใช้ที่มีต่อระบบที่ไม่เป็นการขอข้อมูลจากผู้ใช้โดยตรง เช่น การคลิก การโหวต การแสดงความคิดเห็นบนระบบ เป็นต้น โดย

ตัวอย่างระบบที่ใช้ข้อมูลโดยตรงจากผู้ใช้ เช่น Entrée (Burke, 2000) และ Triplehop's TripMatcher (Kabassi, 2010) ได้ใช้ประโยชน์จากข้อมูลที่สอบถามจากผู้ใช้โดยตรง แต่ปัญหาที่เกิดคือผู้ใช้อาจต้องสูญเสียเวลาในขั้นตอนการตอบคำถามเหล่านี้มากเกินไปซึ่งเป็นการรบกวนผู้ใช้ นอกจากนี้ผู้ใช้ยังอาจไม่สามารถอธิบายลักษณะความพึงพอใจของตนเองได้อย่างถูกต้องผ่านวิธีการนี้ ในขณะที่ PTA (Coyle and Cunningham, 2003) และ GUIDE (Cheverst, et al., 2002) ใช้การได้มาซึ่งข้อมูลความสนใจของผู้ใช้โดยอ้อมผ่านแหล่งที่มาหลายแหล่ง วิธีการนี้สามารถขจัดปัญหาที่เกิดกับระบบที่ขอข้อมูลจากผู้ใช้โดยตรงได้ อย่างไรก็ตามยังมีปัญหาสำคัญคือ ผู้ใช้ที่เริ่มต้นใช้งานระบบ จะยังไม่มีข้อมูลโดยอ้อมของผู้ใช้มากพอที่จะสกัดความสนใจเพื่อแนะนำข้อมูลให้ตรงกับความสนใจได้ ซึ่งเรียกปัญหานี้ว่า โคลด์สตาร์ท เพื่อลดปัญหาระหว่างการได้มาซึ่งข้อมูลความสนใจของผู้ใช้ทั้งสองวิธีข้างต้น จึงมีการผนวกทั้งสองวิธีเข้าด้วยกันเช่นใน TravelPlanner (Chin and Porage, 2001) จะเลือกแบบสอบถามที่มีประโยชน์ที่สุดในการเรียนรู้ความสนใจให้กับผู้ใช้และใช้ข้อมูลอื่นๆ ที่ได้มาจากทางอ้อมด้วย เช่นเดียวกับ Zhu, et al. (2012) ใช้ข้อมูลที่ผู้ใช้ทำการลงทะเบียนเพื่อสร้างแบบจำลองผู้ใช้เริ่มต้น จากนั้นจะใช้ข้อมูลโดยอ้อมผ่านพฤติกรรมผู้ใช้ เช่น พฤติกรรมการคลิก พฤติกรรมการโหวต พฤติกรรมการตอบและพูดคุย เป็นต้น เพื่อเรียนรู้ความสนใจของผู้ใช้ ในทำนองเดียวกันกับ SPETA (García-Crespo, et al., 2009) เมื่อผู้ใช้เริ่มต้นใช้งานระบบ ระบบจะขอข้อมูลเกี่ยวกับประเภทสถานที่ที่ต้องการเข้าไปเยี่ยมชมและการให้คะแนนไปยังสถานที่ที่ผู้ใช้สนใจ จากนั้นก็จะเรียนรู้ความสนใจผ่านพฤติกรรมของผู้ใช้และข้อมูลที่ถูกสกัดจากเครือข่ายสังคมที่ได้เป็นสมาชิกอยู่ จะเห็นว่าการผสมผสานการใช้ข้อมูลทั้งโดยตรงและโดยอ้อมสามารถช่วยลดทั้งปัญหาโคลด์สตาร์ทและการสูญเสียเวลาในการตอบคำถามของผู้ใช้ โดยปัญหาที่เกิดขึ้นในลักษณะการได้มาของข้อมูลด้วยวิธีต่าง ๆ แสดงดังตารางที่ 1

ตาราง 1 ข้อดีของการใช้การได้มาซึ่งข้อมูลความสนใจของผู้ใช้ลักษณะต่าง ๆ

ข้อดี	การได้มาซึ่งข้อมูลความสนใจของผู้ใช้ (User Interest Acquisition)		
	โดยตรง (Explicit)	โดยอ้อม (Implicit)	ผสมผสาน (Hybrid)
1. ผู้ใช้ใช้เวลาน้อยลงในการกรอกข้อมูลความสนใจให้กับระบบ	✗	✓	✓
2. แก้ปัญหาการอธิบายลักษณะหรือระดับความพึงพอใจของผู้ใช้ ในกรณีที่ใช้การขอข้อมูลจากผู้ใช้โดยตรง	✗	✓	✓
3. ระบบมีข้อมูลเพียงพอที่จะนำมาใช้สกัดหาความสนใจ เมื่อผู้ใช้เริ่มต้นใช้ระบบ	✓	✗	✓

ปัญหาโคลด์สตาร์ทเป็นปัญหาที่เกิดขึ้นเมื่อผู้ใช้เริ่มต้นใช้งานระบบแต่ระบบยังไม่มีข้อมูลของผู้ใช้มากพอที่จะสามารถสกัดความสนใจและให้การแนะนำสินค้าหรือบริการได้อย่างถูกต้องโดยเฉพาะอย่างยิ่งในระบบแนะนำข้อมูลเฉพาะบุคคล ซึ่งปัญหาโคลด์สตาร์ทนอกจากจะใช้วิธีการผสมผสานการใช้ข้อมูลจากโดยตรงและโดยอ้อมจากผู้ใช้ที่ได้กล่าวไปข้างต้นแล้ว ยังมีแนวคิดการนำข้อมูลจากผู้ใช้อื่นมาช่วยแก้ปัญหาอีกด้วย เช่น Rashid, et al. (2008) พยายามแก้ไขปัญหานี้โดยใช้ข้อมูลจากผู้ใช้อื่น ๆ ที่มีต่อระบบแนะนำภาพยนตร์ เช่น ข้อมูลภาพยนตร์ที่กำลังเป็นที่นิยม ซึ่งคำนวณจากความถี่ในการเข้าถึงของผู้ใช้ข้อมูล จำนวนของการแสดงความคิดเห็นข้อมูล การให้คะแนนภาพยนตร์ของผู้ใช้ เป็นต้น ภาพยนตร์ที่มีลำดับคะแนนที่สูงก็จะถูกแนะนำให้กับผู้ใช้ที่เริ่มต้นใช้งานระบบ จะเห็นว่าปัญหานี้ถูกแก้ไขในระดับหนึ่งโดยใช้ความสนใจของผู้ใช้อื่นโดยรวม ต่างจาก Hang, et al. (2009) ที่ใช้ข้อมูลจากผู้ใช้อื่นมาพัฒนาอัลกอริทึมที่ใช้ในการวัดค่าความคล้ายระหว่างผู้ใช้และช่วยในการจับคู่ระหว่างผู้ใช้กับสิ่งที่สนใจ โดยมีหลักการว่าผู้เริ่มต้นใช้ระบบจะจับคู่กับสิ่งที่สนใจที่มีค่าสำคัญมากที่สุด (จากผู้ใช้ที่มีความคล้ายคลึงกัน) ซึ่งวิธีการนี้ให้ผลที่น่าพอใจกับระบบที่เน้นไปที่การหาความสนใจรายบุคคลมากกว่าวิธีการแรก

อย่างไรก็ตามยังมีวิธีการที่น่าสนใจที่จะนำมาแก้ไขปัญหานี้อีกวิธีหนึ่งคือการนำข้อมูลมาจากเครือข่ายสังคมของผู้ใช้ ซึ่งวิธีการนี้สามารถใช้ประโยชน์จากทั้งข้อมูลของผู้ใช้เองและผู้ใช้อื่น ๆ รวมถึงไม่เป็นการรบกวนผู้ใช้มากเกินไปในการขอข้อมูลต่าง ๆ เพราะเพียงผู้ใช้ทำการ

อนุญาตให้ระบบเข้าถึงข้อมูล ระบบจะทำการดึงข้อมูลในเครือข่ายเหล่านี้โดยไม่ต้องให้ผู้ใช้กรอกข้อมูลไปยังระบบเอง โดยการนำข้อมูลลักษณะต่าง ๆ ในเครือข่ายสังคมมาใช้ประโยชน์ในการค้นหาความสนใจนั้นจะอธิบายในหัวข้อที่ 3 "เครือข่ายสังคมกับการนำมาใช้ในระบบแนะนำข้อมูลเฉพาะบุคคล"

2.2.2 การเรียนรู้ความสนใจของผู้ใช้ (User Interest Learning)

การเรียนรู้ความสนใจของผู้ใช้เป็นอีกหนึ่งหัวข้อหลักของแบบจำลองผู้ใช้ที่นักวิจัยพยายามพัฒนาเพื่อให้รองรับการเปลี่ยนแปลงความสนใจของผู้ใช้ที่สามารถเกิดขึ้นได้ตลอดเวลาเพื่อให้ระบบแนะนำข้อมูลสามารถแนะนำข้อมูลได้เหมาะสมกับความสนใจของผู้ใช้ในขณะนั้นมากที่สุด โดย Widyantoro, et al. (1999) ได้เสนออัลกอริทึมในการเรียนรู้ความสนใจของผู้ใช้โดยแบ่งความสนใจของผู้ใช้เป็นแบบระยะสั้นเพื่อเรียนรู้ความสนใจในช่วงเวลาสั้น ๆ (Short term) ขณะปัจจุบันและแบบระยะยาว (Long term) เพื่อเรียนรู้ความสนใจโดยรวมจากนั้นจะคำนวณค่าน้ำหนักไปที่แต่ละแบบและเลือกใช้ความสนใจที่มีค่าน้ำหนักมากที่สุด วิธีนี้สามารถเพิ่มความถูกต้องในการค้นหาความสนใจที่เปลี่ยนแปลงไปตามช่วงเวลาได้ และสามารถปรับปรุงความสนใจผู้ใช้ให้ถูกต้องขึ้น สำหรับการเปลี่ยนแปลงความสนใจของผู้ใช้แบบทันทียังไม่ถูกนำมากล่าวถึงมากนัก Zhu, et al. (2012) ได้เสนอการเรียนรู้ความสนใจของผู้ใช้โดยนำข้อมูลโดยอ้อม เช่น พฤติกรรมการคลิก พฤติกรรมการโหวต เป็นต้น จากนั้นคำนวณหาค่าความสนใจในแต่ละเรื่องร่วมกับการคำนวณด้านเวลา (เช่นการกดโหวตเมื่อวานนี้ย่อมมีนัยสำคัญเกี่ยวกับความสนใจมากกว่าการโหวตเมื่อปีที่แล้ว) ค่าความสนใจทั้งหมดของผู้ใช้จะถูกเก็บไว้ในเมทริกซ์และเปลี่ยนแปลงค่าด้วยตัวเองเมื่อผู้ใช้เปลี่ยนแปลงพฤติกรรมต่าง ๆ ข้างต้น ค่าความสนใจที่ได้จากวิธีการนี้สามารถนำไปใช้แนะนำผลิตภัณฑ์หรือบริการได้ตามลำดับค่าความสนใจ นอกจากนี้การเปลี่ยนแปลงค่าความสนใจตามพฤติกรรมในปัจจุบันยังช่วยรองรับการเปลี่ยนแปลงความสนใจในแต่ละช่วงเวลาได้ดียิ่งขึ้น Mezhoudi (2013) ใช้ประโยชน์จากข้อมูลตอบกลับของผู้ใช้ (User Feedback) ปรับปรุงแบบจำลองผู้ใช้ให้กับระบบโดยใช้ข้อมูลตอบกลับจากการปฏิสัมพันธ์ของผู้ใช้ใน 2 แบบ คือ 1) Emoticons Based Feedback ซึ่งเป็นการแสดงระดับความพึงพอใจของผู้ใช้ผ่านทางไอคอนแสดงอารมณ์ ตัวอย่างการนำไปประยุกต์ใช้ เช่น การตอบคำถามข้อเสนอแนะด้วยไอคอนแสดงอารมณ์ในแอพพลิเคชั่นบนโทรศัพท์มือถือ (Arhippainen, et al., 2004) 2) Recommendation Frames เป็นการเก็บข้อมูลจากการปฏิสัมพันธ์โดยการแนะนำ การแนะนำอาจจะแสดงเป็นหน้าต่างป๊อปอัพหรือแสดงบนหน้าเพ็จ ซึ่งใช้กันมากในพาณิชย์อิเล็กทรอนิคส์ (E-Commerce) เพื่อให้คำแนะนำแก่ลูกค้า การนำข้อมูลตอบกลับลักษณะดังกล่าวมาใช้ในการ

เรียนรู้ความสนใจของผู้ใช้เป็นส่วนเสริมในการประเมินความถูกต้องหรือพึงพอใจในการให้คำแนะนำของระบบและยังทำให้ระบบนำข้อมูลเหล่านี้ไปปรับเปลี่ยนความสนใจของผู้ใช้ให้ได้ทันท่วงทีอีกด้วย

จากกระบวนการค้นหาความสนใจของผู้ใช้ทั้งสองส่วนข้างต้น การเลือกใช้วิธีการได้มาซึ่งข้อมูลความสนใจของผู้ใช้โดยตรงนั้น สามารถลดปัญหาโคลด์สตาร์ทที่อาจเกิดขึ้นเมื่อใช้วิธีการได้มาซึ่งข้อมูลความสนใจของผู้ใช้โดยอ้อม อย่างไรก็ตามการขอข้อมูลจากผู้ใช้โดยตรงบางวิธีจะรบกวนผู้ใช้มากเกินไปเพราะผู้ใช้ต้องเสียเวลากับขั้นตอนการกรอกข้อมูลเหล่านี้ ดังนั้นการผสมผสานทั้งสองวิธีการเข้าด้วยกันจึงเป็นอีกทางเลือกหนึ่งที่สามารถลดปัญหาที่เกิดขึ้นจากวิธีการทั้งสองได้ ส่วนการเรียนรู้ความสนใจของผู้ใช้ ซึ่งมีทั้งแนวคิดการแบ่งความสนใจออกเป็นระยะสั้นและระยะยาว การจับพฤติกรรมของผู้ใช้เพื่อปรับเปลี่ยนลักษณะความสนใจ และการใช้ข้อมูลตอบกลับ โดยการใช้ข้อมูลตอบกลับเป็นวิธีการที่มีประสิทธิภาพมากกว่าอีกสองวิธีที่ได้กล่าวไว้ ซึ่งนอกจากจะปรับปรุงความสนใจได้แล้วสามารถนำมาใช้ประเมินความถูกต้องหรือพึงพอใจในการให้คำแนะนำของระบบอีกด้วย

2.3 การสร้างแบบจำลองความสนใจของผู้ใช้ (User Interest Modeling)

การสร้างแบบจำลองผู้ใช้เป็นกระบวนการหลักของระบบเช่นกัน โดยหลังจากที่ระบบได้วิเคราะห์และเรียนรู้ความสนใจจากผู้ใช้แล้ว รูปแบบของการจัดเก็บความสนใจเหล่านี้ก็เป็นสิ่งที่มีความสำคัญ ซึ่งรูปแบบของการจัดเก็บข้อมูลที่นิยมใช้ในปัจจุบัน ดังเช่น

2.3.1 Flat File Database Model

แบบจำลองฐานข้อมูลแบบ Flat File หรือเรียกอีกอย่างหนึ่งว่า System File เป็นการจัดเก็บข้อมูลด้วยข้อความ (Plain Text) ซึ่งโดยมากจะใช้เครื่องหมายจุลภาค (Comma) การเว้นวรรค (Space or Tab) เพื่อแยกข้อมูลแต่ละคอลัมน์ออกจากกัน และใช้การขึ้นบรรทัดใหม่เพื่อแบ่งข้อมูลในแต่ละแถว ดังภาพที่ 2

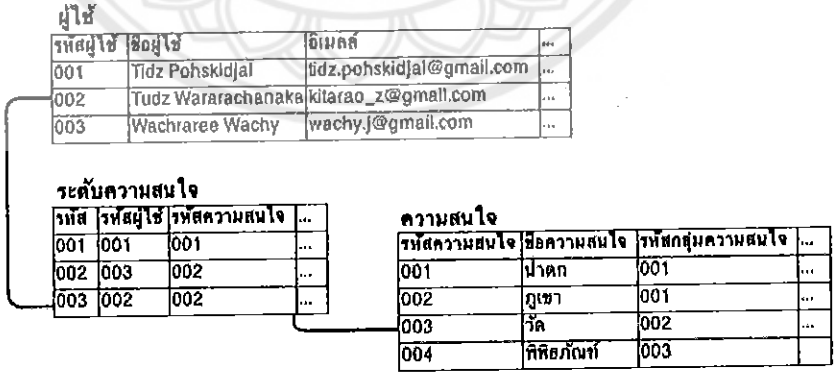
รหัส	ชื่อ	เพศ
1	มงคล	ชาย
2	นาตยา	หญิง
3	ประภาพรรณ	หญิง
4	ธวัช	ชาย
5	ศิริพร	หญิง

ภาพ 2 ตัวอย่างแบบจำลองฐานข้อมูลแบบ Flat File

ข้อจำกัดของการจัดเก็บไฟล์ลักษณะนี้คือ จะไม่สามารถสร้างความสัมพันธ์ระหว่างข้อมูลแต่ละแถวได้ ไม่รองรับการเข้าใช้งานข้อมูลเดียวในเวลาเดียวกัน อาจเกิดความซ้ำซ้อนกันของข้อมูล ทำการค้นหาข้อมูลได้ยาก และแม้จะมีการจัดเก็บข้อมูลเป็นชุดข้อมูล แต่ก็ไม่สามารถนำมาประมวลผลจัดเรียงข้อมูลได้ อย่างไรก็ตาม Flat File เหมาะกับการจัดเก็บข้อมูลบางประเภทที่ต้องการความเร็วในการดึงข้อมูลแต่ไม่มีการเข้าใช้งานพร้อมกัน ซึ่งจะทำความเร็วได้มากกว่าการจัดเก็บด้วยแบบจำลองฐานข้อมูลข้อมูลสัมพันธ์

2.3.2 Relational Database Model

แบบจำลองฐานข้อมูลสัมพันธ์ (Relational Database Model ย่อ RDM) เป็นการจัดเก็บข้อมูลในลักษณะที่เป็นกลุ่มของข้อมูล โดยนิยมเก็บเป็นตาราง (Table) ที่มีความสัมพันธ์กัน เพื่อให้ง่ายต่อการเข้าใจและการประยุกต์ใช้งาน ซึ่งในฐานข้อมูลหนึ่ง ๆ จะมีตารางตั้งแต่ 1 ตารางเป็นต้นไป และในแต่ละตารางนั้นเก็บข้อมูล 2 มิติที่ประกอบด้วยคอลัมน์ (Column) และแถว (Row) ดังภาพที่ 3



ภาพ 3 ตัวอย่างแบบจำลองฐานข้อมูลสัมพันธ์

ข้อดีของการจัดเก็บข้อมูลลักษณะนี้คือสามารถสร้างความสัมพันธ์ระหว่างข้อมูลได้ ลดความซ้ำซ้อนกันของการเก็บข้อมูล ค้นหาหรือสอบถามข้อมูลได้ง่าย นอกจากนี้การใช้ RDBMS (Relational Database Model Management System) ยังช่วยในเรื่องความปลอดภัยในการเข้าถึงข้อมูลโดยการจัดลำดับสิทธิ์ (Permission Level) รองรับการเข้าใช้งานข้อมูลเดียวในเวลาเดียวกัน รวมถึงมีการสำรองข้อมูล (Backup) อย่างไรก็ตามในกลุ่มข้อมูลที่มีความสัมพันธ์จำนวนมาก ซึ่งต้องมีการเชื่อมโยงกันในหลายตาราง ทำให้อาจเสียเวลาในการดึงข้อมูลมากขึ้น

2.3.3 Ontology-based Model

ออนโทโลจี (ontology) เป็นเทคนิคหนึ่งในการสร้างองค์ความรู้ (Knowledge-based) โดยถูกนำมาใช้มากขึ้นในหลายสาขาทางวิทยาการคอมพิวเตอร์ เนื่องจากมีประสิทธิภาพในการวิเคราะห์ข้อมูล (Castells, et al., 2007) โดยปัจจุบันเริ่มมีการนำออนโทโลจีมาใช้เป็นรูปแบบการจัดเก็บข้อมูล เช่น การจัดเก็บข้อมูลการเดินทางและสถานที่ท่องเที่ยว (นฤพนธ์ พนาวงศ์ และจักรกฤษณ์ เสน่ห์, 2553) เพื่อให้เกิดความเชื่อมโยงระหว่างข้อมูล ซึ่งการจัดเก็บจะต้องอาศัยการจัดเก็บในภาษาที่มีความสามารถในการอธิบายถึงความหมายและความสัมพันธ์ระหว่างข้อมูล โดยส่วนใหญ่เป็นภาษาที่มีพื้นฐานมาจากภาษา XML เช่น RDF RDF-S OWL เป็นต้น ข้อดีของการจัดเก็บข้อมูลลักษณะนี้คือ ช่วยในเรื่องการเข้าใจความหมายของข้อมูล ลดความกำกวมในความหมายของข้อมูลได้ (ไกรศักดิ์ เกษร, 2555) นอกจากนี้ยังสามารถค้นหาข้อมูลได้ง่ายเพราะมีภาษาในการสืบค้นเช่นเดียวกับแบบจำลองฐานข้อมูลสัมพันธ์ แต่ปัญหาสำคัญของออนโทโลจีคือ ยังไม่มีความยืดหยุ่นและมีโครงสร้างที่ไม่เปลี่ยนแปลง ส่งผลให้เมื่อนำไปใช้จัดเก็บข้อมูลที่ต้องมีการเปลี่ยนแปลงจะทำได้ยากกว่าแบบจำลองฐานข้อมูลสัมพันธ์

จากรูปแบบของการจัดเก็บข้อมูลข้างต้น พบว่าการจัดเก็บข้อมูลแบบ Flat File ไม่เหมาะ สมที่จะนำมาจัดเก็บข้อมูลแบบจำลองความสนใจของผู้ใช้ เนื่องจากไม่สามารถสร้างความสัมพันธ์ระหว่างข้อมูลได้ การค้นหาและการเปลี่ยนแปลงข้อมูลทำได้ยาก และไม่มีความปลอดภัยในข้อมูล ส่วนการจัดเก็บแบบ Ontology-based แม้จะได้รับความสนใจมากขึ้นเรื่อย ๆ เพราะสามารถอธิบายความหมายของข้อมูลได้ สืบค้นข้อมูลได้ โดยนิยมใช้จัดเก็บข้อมูลที่มีลักษณะเป็นโดเมน เช่น ข้อมูลสถานที่ท่องเที่ยว เป็นต้น แต่การนำมาใช้ในการจัดเก็บข้อมูลที่ต้องมีการเปลี่ยนแปลง เช่น ความสนใจของผู้ใช้ที่ต้องมีการปรับเปลี่ยนบ่อยครั้ง การปรับเปลี่ยนจะทำได้ยาก และยังไม่มีระบบการจัดการ (Management System) มากนัก ทำให้อาจมีปัญหาเรื่องความปลอดภัยในข้อมูลได้ด้วย ส่วนการจัดเก็บแบบฐานข้อมูลสัมพันธ์นั้นเป็นรูปแบบการจัดเก็บที่ได้รับความนิยมในปัจจุบันรวมถึงมีความเหมาะสมในการจัดเก็บข้อมูลความสนใจของผู้ใช้

เนื่องจากสามารถเก็บข้อมูลที่มีความสัมพันธ์กัน สามารถสร้าง ค้นหา และเปลี่ยนแปลงข้อมูลได้ง่าย รวมถึงมีระบบการจัดการหลายระบบให้เลือกใช้ซึ่งสามารถเพิ่มความปลอดภัยในข้อมูลได้ โดยผู้วิจัยได้สรุปปัญหาหลักที่เกิด ดังตารางที่ 2

ตาราง 2 ข้อดีของการสร้างแบบจำลองความสนใจของผู้ใช้ด้วยรูปแบบการจัดเก็บต่างกัน

ข้อดี	การสร้างแบบจำลองความสนใจของผู้ใช้ (User Interest Modeling)		
	Flat File	Relational	Ontology-based
1. สามารถสร้างความสัมพันธ์ระหว่างข้อมูลได้	✗	✓	✓
2. การค้นหาข้อมูลทำได้ง่าย	✗	✓	✓
3. การเปลี่ยนแปลงข้อมูลทำได้ง่าย	✗	✓	✗
4. มีความปลอดภัยของข้อมูล	✗	✓	✗

2.4 เทคนิคการแนะนำข้อมูลให้ผู้ใช้ (Recommendation Techniques)

การแนะนำข้อมูลเป็นกระบวนการที่ช่วยค้นหาสินค้าหรือบริการที่ผู้ใช้กำลังสนใจ เปรียบกับการจับคู่สิ่งที่เหมาะสมกับความต้องการของผู้ใช้นั่นเอง ซึ่งเทคนิคที่ใช้ในการแนะนำข้อมูลมีอยู่หลายวิธีการ เช่น Content-based, Collaborative, Demographic, Knowledge-based และวิธีแบบผสม (Hybrid) เป็นต้น แต่วิธีที่นิยมนำมาใช้มากที่สุด คือวิธีการ Content-based และ Collaborative โดยวิธีการแบบ Content-based นั้นเป็นการแนะนำโดยการวิเคราะห์จากพฤติกรรมในอดีตของผู้ใช้เอง ในขณะที่วิธีการแบบ Collaborative เป็นการแนะนำโดยใช้ข้อมูลจากคำแนะนำที่ให้กับผู้ใช้อื่น

จากเทคนิคในการแนะนำข้อมูลข้างต้น ปัญหาหลักที่เกิดขึ้นคือ การใช้เทคนิคแบบ Content-based ในกรณีผู้ใช้ใช้งานระบบครั้งแรก แล้วระบบไม่มีข้อมูลความสนใจของผู้ใช้ จะทำให้ระบบไม่สามารถแนะนำข้อมูลที่ถูกต้องให้กับผู้ใช้ได้ (เกิดปัญหาโคลด์สตาร์ท) แต่เทคนิคแบบ Collaborative ระบบจะพยายามเชื่อมต่อสิ่งที่ผู้ใช้และผู้ใช้อื่นน่าจะมีความสนใจเหมือนกัน และแนะนำสิ่งที่เคยแนะนำให้ผู้ใช้อื่นให้กับผู้ใช้ ทำให้ลดปัญหาโคลด์สตาร์ทไปได้ อย่างไรก็ตามวิธีการแบบ Collaborative มีข้อจำกัดคือในความเป็นจริงผู้ใช้แต่ละคน อาจจะมีความสนใจพิเศษเฉพาะ ทำให้ระบบอาจจะไม่ได้แนะนำข้อมูลที่ควรจะต้องแนะนำให้กับผู้ใช้ ซึ่งตรงกันข้ามกับแบบ Content-based ที่จะไม่เกิดปัญหานี้ เนื่องจากใช้ข้อมูลจากผู้ใช้เองเป็นหลัก

บทที่ 3
วิธีดำเนินการวิจัย

ในบทนี้จะอธิบายถึงวิธีดำเนินการวิจัยโดยประกอบด้วยส่วนต่างๆ ดังต่อไปนี้คือ ประชากรและกลุ่มตัวอย่าง การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล วิธีการดำเนินงานวิจัย เครื่องมือที่ใช้ในการวิจัย และกรอบแนวคิดในการออกแบบและพัฒนาระบบ ซึ่งจะอธิบายดังต่อไปนี้

3.1 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยประกอบด้วย เครื่องมือที่ใช้ในการวิจัยเชิงคุณภาพ เครื่องมือที่ใช้ในการพัฒนาระบบ และเครื่องมือที่ใช้ในการประเมินระบบ ดังนี้

1. เครื่องมือที่ใช้ในการวิจัยเชิงคุณภาพ ได้แก่ แบบสอบถามความต้องการเกี่ยวกับการท่องเที่ยวของผู้ใช้ โดยเก็บรวบรวมข้อมูลรูปแบบความต้องการเพื่อใช้ในการพัฒนาระบบ
2. เครื่องมือที่ใช้ในการพัฒนาระบบ แบ่งออกเป็น 2 ประเภท ดังนี้

2.1 ด้านโปรแกรมคอมพิวเตอร์หรือซอฟแวร์ (Software)

2.1.1 PHP 5.3.28 ใช้เป็นภาษาหลักในการพัฒนาระบบ

2.1.2 PostgreSQL 8.4.2.1 ใช้ในการจัดการฐานข้อมูล

2.1.3 pgAdminIII 1.12.3 ใช้เป็นเครื่องมือช่วยติดต่อฐานข้อมูล PostgreSQL

2.1.4 Apache 2.2.26 ใช้ในการจำลองเว็บเซิร์ฟเวอร์

2.1.5 AJAX, Jquery และ JavaScipt ใช้สนับสนุนการแสดงและส่งค่าการทำงานต่าง ๆ ร่วมกับ PHP เพื่อให้แสดงผลได้รวดเร็ว หลากหลายรูปแบบ และสวยงามมากขึ้น

2.1.6 Facebook API 2.0 ใช้เป็นเครื่องมือในการติดต่อและจัดการข้อมูลจากเฟสบุ๊ค

2.2 ด้านอุปกรณ์คอมพิวเตอร์หรือฮาร์ดแวร์ (Hardware)

2.2.1 หน่วยประมวลผล (CPU) Intel Core i5-3210M 2.5 GHz

2.2.2 หน่วยความจำหลัก (Ram) DDR34 GB

2.3 ระบบปฏิบัติการ (OS)

2.3.1 Microsoft Windows 7 64-bit

3. เครื่องมือที่ใช้ในการประเมินระบบ คือ แบบประเมินประสิทธิภาพของระบบและแบบสอบถามความพึงพอใจของผู้ใช้งาน

3.2 การเก็บรวบรวมข้อมูล

1. การเก็บรวบรวมข้อมูลสถานที่ท่องเที่ยวในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้ข้อมูลสถานที่ท่องเที่ยว 5 จังหวัด ได้แก่ พิษณุโลก อุตรดิตถ์ ตาก สุโขทัย และ เพชรบูรณ์ โดย Harfield (2014) ร่วมกับข้อมูลสถานที่ท่องเที่ยวจากการท่องเที่ยวแห่งประเทศไทย (ททท.) เป็นต้นแบบพื้นที่การแนะนำของระบบ โดยเก็บเป็นฐานข้อมูลสถานที่ท่องเที่ยว จำนวน 223 แห่ง แบ่งเป็น 6 หมวดหมู่ ดังนี้ สถานที่ท่องเที่ยวทางวัฒนธรรม 20 แห่ง สถานที่ท่องเที่ยวทางประวัติศาสตร์ 65 แห่ง สถานที่ท่องเที่ยวทางธรรมชาติ 70 แห่ง สถานที่ท่องเที่ยวเพื่อความบันเทิง 20 แห่ง สถานที่ท่องเที่ยวเพื่อการศึกษา 33 แห่ง และสถานที่ท่องเที่ยวทางวิถีชีวิตชุมชน 16 แห่ง

2. การเก็บรวบรวมความต้องการของผู้ใช้ โดยเก็บรวบรวมข้อมูลรูปแบบความต้องการเพื่อใช้ในการพัฒนาระบบ ผู้วิจัยใช้แบบสอบถามออนไลน์ (Online Questionnaire) โดยมีจำนวนผู้ตอบแบบสอบถามทั้งสิ้นเป็นจำนวน 450 คน (รายละเอียดของผู้ตอบแบบสอบถาม อธิบายเพิ่มเติมในบทที่ 4 ผลการวิจัย)

3. การเก็บรวบรวมข้อมูลจากการประเมินประสิทธิภาพของระบบ เพื่อประเมินประสิทธิภาพของระบบว่าสอดคล้องกับวัตถุประสงค์และสมมติฐานของการวิจัยเพียงใด โดยใช้กลุ่มตัวอย่างจำนวน 42 คน และประเมินความพึงพอใจของระบบ เพื่อประเมินผลการตอบสนองความต้องการของผู้ใช้บริการหรือนักท่องเที่ยวมากน้อยเพียงใด โดยใช้ผู้ตอบแบบสอบถามจำนวน 91 คน (รายละเอียดของกลุ่มตัวอย่าง อธิบายเพิ่มเติมในบทที่ 4 ผลการวิจัย)

3.3 วิธีการดำเนินงานวิจัย

งานวิจัยนี้ มีขั้นตอนของการดำเนินการวิจัย ดังนี้

1. การศึกษาเอกสาร ข้อมูล และงานวิจัยที่เกี่ยวข้อง โดยวิเคราะห์หลักการที่นำไปใช้ในการพัฒนาระบบ ดังนี้ (รายละเอียดเพิ่มเติมในบทที่ 2)

1.1 ระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล

1.2 องค์ประกอบของระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล

1.3 เครือข่ายสังคมกับการนำมาใช้ในระบบแนะนำข้อมูลเฉพาะบุคคล

1.4 เทคนิคตรวจสอบการเปลี่ยนแปลงความสนใจของผู้ใช้ (ข้อมูลตอบกลับของผู้ใช้)

2. วิเคราะห์ ออกแบบ และพัฒนาระบบ

2.1 เก็บรวบรวมข้อมูลสถานที่ท่องเที่ยวใน 5 จังหวัดต้นแบบ

2.2 เก็บรวบรวมความต้องการและวิเคราะห์ความต้องการของผู้ใช้

2.3 ออกแบบกรอบการทำงานจากความต้องการของผู้ใช้

2.4 เลือกเครื่องมือที่จะนำมาใช้พัฒนาระบบ

2.5 พัฒนาระบบแนะนำข้อมูลการท่องเที่ยวเฉพาะบุคคล

3. ทดสอบการใช้งานระบบ ปรับปรุงระบบ และตรวจสอบความถูกต้อง

4. ทดสอบประสิทธิภาพระบบตามสมมติฐานที่ตั้งไว้

5. ติดตั้งระบบต้นแบบ เว็บแอปพลิเคชัน PTIS บนเว็บไซต์จริง เพื่อให้ผู้ที่สนใจทดลองใช้

6. ประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบ

7. สรุปผล อภิปรายผล และข้อเสนอแนะ รวมทั้งจัดทำรูปเล่มรายงาน

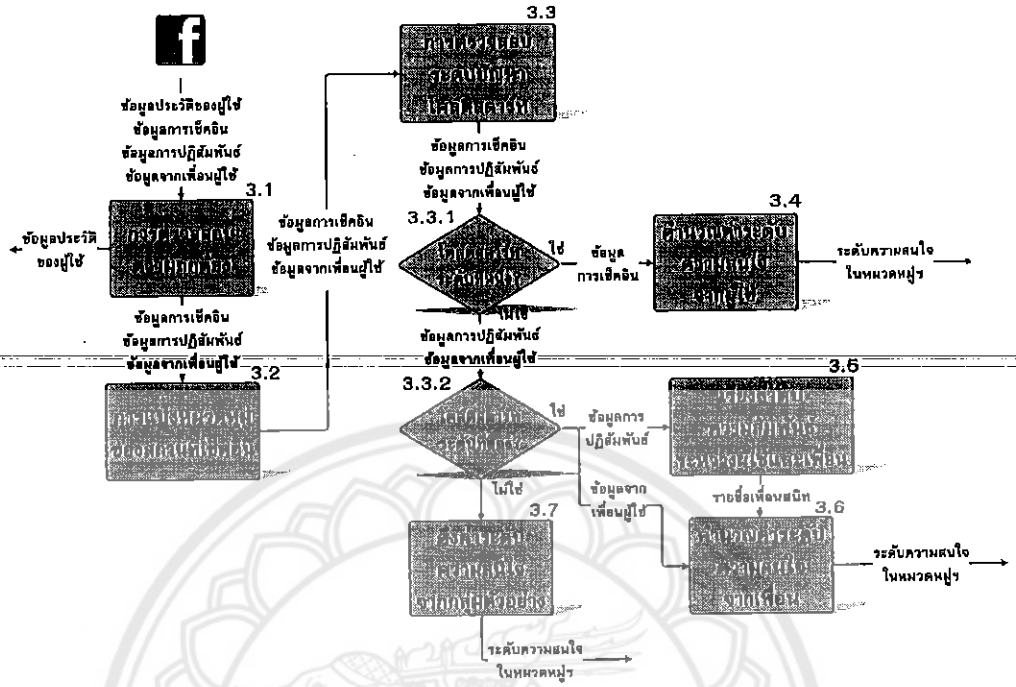
3.4 กรอบแนวคิดการวิจัย

ในการวิเคราะห์ความสนใจของผู้ใช้ดังภาพที่ 4 ในขั้นตอนนี้ ใช้การได้มาซึ่งข้อมูลความสนใจจากผู้ใช้โดยตรง (Explicit) โดยขอใช้ข้อมูลจากเครือข่ายสังคมของผู้ใช้ เพื่อแก้ไขปัญหาโคล์ดสตาร์ทและลดขั้นตอนที่ยุ่งยากในการให้ข้อมูลความสนใจ นอกจากนี้ในการศึกษาพฤติกรรมการใช้งานบนเครือข่ายสังคมเฟสบุ๊ค พบว่าส่วนหนึ่งของผู้ใช้ที่เข้าใช้งานเฟสบุ๊คเป็นประจำ ไม่ได้ทำเช็คอินในทุกครั้งที่ใช้งาน ซึ่งอาจทำให้ผู้ใช้ไม่มีข้อมูลประวัติการเช็คอินที่มากเพียงพอในการบอกถึงความสนใจ ดังนั้นผู้วิจัยจึงได้แบ่งปัญหาโคลด์สตาร์ทที่เกิดกับผู้ใช้ที่เริ่มต้นใช้งานระบบเป็น 2 ระดับ คือ

1. โคลด์สตาร์ทระดับที่ 1 กรณีที่ระบบไม่มีข้อมูลความสนใจของผู้ใช้ที่เริ่มใช้งานระบบ จะใช้ข้อมูลการเช็คอินในอดีตของผู้ใช้ในการวิเคราะห์ความสนใจ

2. โคลด์สตาร์ทระดับที่ 2 กรณีที่ไม่สามารถใช้ข้อมูลการเช็คอินของผู้ใช้ในโคลด์สตาร์ทในระดับที่ 1 ได้ จะใช้ข้อมูลการเช็คอินจากเพื่อนที่มีพฤติกรรมหรือความสนใจคล้ายกัน

โดยหากไม่สามารถแก้ไขโคล์ดสตาร์ทในระดับที่ 2 ได้ จะใช้ค่าพื้นฐาน (Default) ที่เป็นค่าเฉลี่ยจากการคำนวณในกลุ่มตัวอย่าง



ภาพ 4 วิธีการวิเคราะห์ข้อมูลจากเครือข่ายสังคม

ในการนำข้อมูลจากเครือข่ายสังคมมาใช้จะใช้การดึงข้อมูลมาจากเฟสบุ๊คผ่าน Facebook API ซึ่งจะได้ข้อมูลเป็นแบบ JSON โดยข้อมูลที่สามารถดึงมาจากเฟสบุ๊คประกอบด้วยหลายฟิลด์ (Field) ข้อมูล ดังภาพที่ 5

fields

about	languages
address	last_name
age_range	link
bio	locale
birthday	location
can_send_to_mobile	meeting_for
cover	middle_name
currency	name
devices	name_format
education	payment_engagement
email	payment_mobile_pricepoints
favorite_athletes	payment_pricepoints
favorite_teams	payment_test_group
first_name	political
gender	quotes
	relationship_status

ภาพ 5 ตัวอย่างฟิลด์ข้อมูลของผู้ใช้ที่สามารถดึงมาจากเฟสบุ๊ค

แต่อย่างไรก็ตามระบบจะเลือกใช้เฉพาะฟิลด์ที่มีประโยชน์ต่อการระบุตัวตนและวิเคราะห์ความสนใจของผู้ใช้ โดยฟิลด์ข้อมูลที่ผู้วิจัยนำมาใช้ภายในระบบประกอบด้วย

1. ข้อมูลเพื่อระบุตัวตนของผู้ใช้ แสดงดังตารางที่ 4
   1.1 user id (รหัสผู้ใช้) โดยใช้ฟิลด์ id (หมายเลขผู้ใช้ของเฟสบุ๊ค)
   1.2 username (ชื่อผู้ใช้) โดยใช้ฟิลด์ first_name(ชื่อ) last_name (นามสกุล)
   1.3 email (อิเมล์) โดยใช้ฟิลด์ email (อิเมล์)

โดยข้อมูลเพื่อระบุตัวตนของผู้ใช้เหล่านี้เป็นข้อมูลประวัติของผู้ใช้ที่จะส่งไปยังกระบวนการ 2 เพื่อนำข้อมูลประวัติผู้ใช้บนเครือข่ายสังคมไปสร้างหรือปรับปรุงข้อมูลประวัติผู้ใช้ในแบบจำลอง ซึ่งได้กล่าวมาแล้วก่อนหน้านี้

2. ข้อมูลการเช็คอินในแต่ละสถานที่ นำมาจากฟิลด์ location โดยแสดงรายละเอียดดังตารางที่ 5 และข้อมูลรายละเอียดของสถานที่ที่ถูกเช็คอินดังตารางที่ 6 เพื่อนำมาใช้ในการวิเคราะห์ความสนใจของผู้ใช้

ตาราง 4 ข้อมูลเพื่อระบุตัวตนของผู้ใช้

ข้อมูลที่ได้จากเฟสบุ๊ค	ความหมาย
{ "id": "100000965654095",	หมายเลขผู้ใช้ของเฟสบุ๊ค ใช้ระบุตัวตนในการเข้าถึงระบบ
"first_name": "TiDz", "last_name": "PohSkidJai",	ชื่อและนามสกุล ใช้แสดงบนส่วนติดต่อกับผู้ใช้ (UI)
"email": "tidz.pohskidjai@gmail.com"}	อิเมลล์ ใช้สำหรับแจ้งข้อมูลต่าง ๆ ของระบบ

ตาราง 5 ข้อมูลการเช็คอินของผู้ใช้จำนวน 1 เช็คอิน

ข้อมูลที่ได้จากเฟสบุ๊ค	ความหมาย
{"place": {	การเช็คอินในสถานที่จำนวน 1 เช็คอิน
"id": "196230743801741",	หมายเลขไอดีของสถานที่ ที่ทำการเช็คอิน
"name": "วัดพระศรีรัตนมหาธาตุราชวรวิหาร",	ชื่อของสถานที่ ที่ทำการเช็คอิน
"location": {	ตำแหน่งที่ตั้งของสถานที่ ที่ทำการเช็คอิน

ข้อมูลที่ได้จากเฟสบุ๊ค	ความหมาย
"latitude": 17.429238993972,	ละติจูดของสถานที่ ที่ทำการเช็คอิน
"longitude": 99.811871548508,	ลองจิจูดของสถานที่ ที่ทำการเช็คอิน
"street": "อ.ศรีสัชนาลัย" }	ข้อมูลที่อยู่ของสถานที่ ที่ทำการเช็คอิน
}, "id": "750723874966970",	หมายเลขไอดีของโพสที่ทำการเช็คอิน
"created_time": "2013-04-13T09:21:03+0000"}	เวลาในการสร้างโพสที่ทำการเช็คอิน

ตาราง 6 หมวดหมู่ (category_list) ของสถานที่ที่ถูกเช็คอิน

ข้อมูลที่ได้จากเฟสบุ๊ค	ความหมาย
{ "id": "196230743801741",	หมายเลขไอดีของสถานที่ ที่ทำการเช็คอิน
"name": "วัดพระศรีรัตนมหาธาตุราชวรวิหาร",	ชื่อของสถานที่ ที่ทำการเช็คอิน
"category_list": [	รายการหมวดหมู่ของสถานที่ที่จัดแบ่งบนเฟสบุ๊ค
{ "id": "124861974254366",	หมายเลขไอดีหมวดหมู่ที่ 1 ของสถานที่
"name": "Tours & Sightseeing" },	ชื่อหมวดหมู่ที่ 1 ของสถานที่
{ "id": "197097220301977",	หมายเลขไอดีหมวดหมู่ที่ 2 ของสถานที่
"name": "Historical Place"}] }	ชื่อหมวดหมู่ที่ 2 ของสถานที่

โดยข้อมูลที่นำมาใช้ในการวิเคราะห์ความสนใจ คือ 1) Place -> id (รหัสสถานที่) 2) created_time (เวลาที่เช็คอิน) 3) category_list (หมวดหมู่ของสถานที่) เป็นต้น

3.4.1 การตรวจสอบความถูกต้อง (Cleansing)

เมื่อได้ข้อมูลของผู้ใช้ตามฟิลด์ข้อมูลที่กล่าวข้างต้นแล้ว ผู้วิจัยจะตรวจสอบความถูกต้องของข้อมูลการเช็คอินโดยเลือกใช้ข้อมูลการเช็คอินที่มีจุดประสงค์เพื่อการท่องเที่ยวเท่านั้น เนื่องจากการศึกษาพบว่าผู้ใช้ในเครือข่ายสังคมมักจะเช็คอินในสถานที่หนึ่ง ๆ เป็นจำนวนมากเหมือน ๆ กัน เช่น ห้างสรรพสินค้า โรงภาพยนตร์ บ้านของผู้ใช้ เป็นต้น ซึ่งสถานที่นี้ส่วนมากเป็นกิจกรรมในการดำเนินชีวิตทั่วไป ซึ่งหากวิเคราะห์ความสนใจจากข้อมูลที่ไม่ผ่านการตรวจสอบความถูกต้องเหล่านี้ ความสนใจที่วิเคราะห์ได้อาจไม่สอดคล้องกับความสนใจของผู้ใช้ในสถานที่ท่องเที่ยวอย่างแท้จริง อย่างไรก็ตามมีบางกรณีที่สถานที่เหล่านี้อาจจัดว่าเป็นสถานที่ท่องเที่ยวได้ ดังนั้นผู้วิจัยจึงใช้วิธีการพิจารณามิติของเวลาเข้ามาแทนการตัดสถานที่ที่ไม่ใช่จุดประสงค์เพื่อการ

ท่องเที่ยวออกไปทันที โดยจะเลือกตัดสถานที่มีการเช็คอินมากกว่าหนึ่งครั้งต่อสัปดาห์ อีกทั้งยังใช้ค่าของเวลาลดความซ้ำซ้อนของการเช็คอินที่อาจเกิดขึ้นหลายครั้งในวันเดียวกัน ยกตัวอย่างเช่น การโพสรูปภาพหลาย ๆ รูปภาพที่ติดกับเช็คอินในสถานที่เดียวกันในวันเดียวกัน สถานที่ที่ซ้ำซ้อนเหล่านี้จะถูกตัดออกให้เหลือเพียงหนึ่งครั้งในหนึ่งวัน

3.4.2 การแบ่งหมวดหมู่ของสถานที่เช็คอิน (Check-in Classification)

การแบ่งหมวดหมู่ของสถานที่เช็คอินจะเริ่มกระบวนการเมื่อการเช็คอินผ่านการตรวจสอบความถูกต้องเรียบร้อยแล้ว โดยการจับคู่ category_list ที่ได้จากการเช็คอินของผู้ใช้กับฐานข้อมูล category_list ที่ผู้วิจัยได้จัดทำขึ้น

ในการสร้างฐานข้อมูล category_list นั้น ใช้วิธีการแบ่งหมวดหมู่จากคำสำคัญที่เกี่ยวข้องกับหมวดหมู่นั้น อย่างไรก็ตามพบว่าเฟสบุ๊คมีฟังก์ชันให้ผู้ใช้สามารถสร้าง category_list เองได้ ทำให้เกิดกรณีชื่อ category_list ต่างกัน แต่หมายถึงสิ่งเดียวกัน หรือความเฉพาะเจาะจงของชื่อ ซึ่งในระบบต้นแบบนี้ผู้วิจัยได้จัดการกับปัญหาข้างต้นโดยวิธีการแบบ Knowledge-based เนื่องจากการพัฒนาเทคนิคเพื่อตรวจความซ้ำซ้อนของหมวดหมู่สถานที่ท่องเที่ยวมีความซับซ้อนและไม่อยู่ในขอบเขตของงานวิจัยนี้ ดังนั้นวิธีการเลือก category_list ในการทำฐานข้อมูลเพื่อจำแนกหมวดหมู่ ทำได้โดยใช้ category_list พื้นฐานของเฟสบุ๊ค category_list ที่เก็บจากกลุ่มตัวอย่าง และ category_list จากผู้ใช้เฟสบุ๊คทั่วไปที่ตั้งสถานะการเช็คอินเป็นแบบ public ซึ่งฐานข้อมูล category_list จะแบ่งตามกลุ่มความสนใจในสถานที่ท่องเที่ยวออกเป็น 6 หมวดหมู่ (รายละเอียดเพิ่มเติมในภาคผนวก จ) โดยปรับปรุงมาจากหมวดหมู่หลักของการท่องเที่ยวแห่งประเทศไทย (tourismthailand.org) ได้แก่

1. สถานที่ท่องเที่ยวทางวัฒนธรรม หมายถึงแหล่งท่องเที่ยวที่ทำให้นักท่องเที่ยวได้รู้จักวิถีชีวิต วัฒนธรรม และความเป็นอยู่ของผู้คนในชุมชน ในสถานที่ที่จัดนิทรรศการซึ่งแสดงเรื่องราวและงานศิลปหัตถกรรมของผู้คนในชุมชน เช่น งานประเพณี การแสดงศิลปวัฒนธรรม งานแสดงสินค้าพื้นเมือง เป็นต้น โดยจัด category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ ยกตัวอย่างเช่น Cultural_Gifts_Store, Fairground, Shopping District เป็นต้น

2. สถานที่ท่องเที่ยวทางประวัติศาสตร์ หมายถึงสถานที่ท่องเที่ยวที่แสดงถึงร่องรอยของความรุ่งเรืองในอดีตของสถานที่สำคัญหลายแห่ง ซึ่งบางแห่งได้รับคัดเลือกให้เป็นแหล่งมรดกโลกอีกด้วย เช่น โบราณสถาน อุทยานประวัติศาสตร์ วัด ศาสนสถาน สิ่งก่อสร้างที่มีคุณค่าทางศิลปะและสถาปัตยกรรม เป็นต้น โดยจัด category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ ยกตัวอย่างเช่น Historical_Place, Buddhist_Temple, Religious_Organization เป็นต้น

3. สถานที่ท่องเที่ยวเพื่อความบันเทิง หมายถึงสถานที่ท่องเที่ยวเชิงสันทนาการและบันเทิง ซึ่งมีหลายรูปแบบ เช่น สวนสัตว์ สวนสนุก ย่านบันเทิง สวนสาธารณะ สิ่งปลูกสร้างต่าง ๆ เป็นต้น โดยจัด category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ ยกตัวอย่างเช่น Market, Park, Art&Entertaimment, Outdoor, Zoo เป็นต้น

4. สถานที่ท่องเที่ยวทางธรรมชาติ หมายถึงสถานที่ท่องเที่ยวที่เกิดจากแหล่งธรรมชาติหลากหลายประเภท ตั้งแต่ ภูเขา ผืนป่า น้ำตก โถงถ้ำ ทะเลสาบ ทุ่งดอกไม้ น้ำพุร้อน ฯลฯ โดยผู้วิจัยได้เลือก category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ยกตัวอย่างเช่น National_Park, Mountain, River, Rock Climbing, Lake เป็นต้น

5. สถานที่ท่องเที่ยวเพื่อการศึกษา หมายถึงสถานที่ท่องเที่ยวเชิงวิชาการสำหรับผู้ที่สนใจใฝ่หาความรู้ อาทิ ห้องสมุด พิพิธภัณฑ์เพื่อการศึกษา ศูนย์ฝึกอบรม เป็นต้น โดยจัด category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ ยกตัวอย่างเช่น Museum, Education เป็นต้น

6. สถานที่ท่องเที่ยวทางวิถีชีวิตชุมชน เป็นสถานที่ท่องเที่ยวที่แสดงถึงวิถีชีวิตของคนในชุมชน เช่น การท่องเที่ยวในรูปแบบโฮมสเตย์ หรือจะเป็นตลาด ตลาดน้ำ ที่เป็นแหล่งท่องเที่ยวที่เต็มไปด้วยสีสันของวิถีชีวิตไทยดั้งเดิมและอุดมด้วยอาหารไทย หลากหลาย โดยจัด category_list ในเฟสบุ๊คให้อยู่ในหมวดหมู่นี้ยกตัวอย่างเช่น Region, Farmers Market เป็นต้น

เมื่อได้ข้อมูลการเช็คอินที่จัดลงหมวดหมู่เรียบร้อยแล้ว จะเริ่มขั้นตอนการวิเคราะห์ความสนใจ โดยใช้ข้อมูลจากการแบ่งหมวดหมู่ของสถานที่เช็คอินในขั้นตอนก่อนหน้า คำนวณหาค่าระดับความสนใจ (Interest Level - I) ในแต่ละหมวดหมู่ของการท่องเที่ยว ซึ่งใช้เทคนิคการแสดงค่าความสนใจด้วยตัวเลข (Numerical Representations) (Moreno, et al., 2013) ซึ่งลักษณะของเทคนิคนี้ ค่าระดับความสนใจ จะถูกคำนวณไปยังแต่ละหมวดหมู่ของการท่องเที่ยว โดยให้ค่าระดับความสนใจมีค่า 0.0 ถึง 1.0 โดย 0 หมายถึงไม่มีความสนใจ และ 1 หมายถึงมีความสนใจระดับสูงสุด

3.4.3 การตรวจสอบระดับปัญหาโคล์ดสตาร์ท (Cold Start Level Detection)

จากนั้นระบบจะตรวจสอบเป็นว่าผู้ใช้เกิดปัญหาโคลด์สตาร์ทระดับใด โดยพิจารณาจากข้อมูลเช็คอินว่าเพียงพอที่จะนำมาวิเคราะห์ความสนใจหรือไม่ โดยหากข้อมูลการเช็คอินที่ผ่านการคลีนซิงแล้ว มีไม่ต่ำกว่า 10 เช็คอิน จะถือว่าเป็นปัญหาโคล์ดสตาร์ทระดับที่ 1 (จากผลการทดสอบที่ 1) ระบบจะใช้ข้อมูลการเช็คอินของผู้ใช้เอง หากมีน้อยกว่า จะเป็นโคล์ดสตาร์ทระดับที่ 2 จะใช้ข้อมูลการเช็คอินของกลุ่มเพื่อนสนิท หากข้อมูลจากกลุ่มเพื่อนไม่เพียงพออีกก็จะใช้ข้อมูลจากกลุ่มตัวอย่าง ดังนี้

1) โคลด์สตาร์ทระดับที่ 1 ระบบจะตรวจสอบว่าผู้ใช้นั้นมีข้อมูลการเช็คอินเพียงพอที่จะนำมาวิเคราะห์ความสนใจหรือไม่ โดยหากข้อมูลการเช็คอินที่ผ่านการตรวจสอบความถูกต้องแล้ว มีไม่ต่ำกว่า 10 เช็คอิน ซึ่งระบบถือว่ามีข้อมูลเพียงพอ (จากการทดลองที่ 1) ก็จะใช้ข้อมูลของผู้ใช้เองในการวิเคราะห์ความสนใจ (3.4) โดยวิธีการคำนวณค่าระดับความสนใจ จะใช้ข้อมูลประวัติสถานที่ที่ผู้ใช้ได้เคยไปก่อนหน้าผ่านการเช็คอินในเฟสบุ๊ค โดยนับจำนวนของการเช็คอินที่ปรากฏในแต่ละหมวดหมู่การท่องเที่ยวของแต่ละผู้ใช้ คำนวณดังสมการที่ 1

$$I(c) = \frac{n_c}{\sum_{k=1}^{6} n_k}, \quad 1 \le c \le 6 \qquad (1)$$

$I(c)$ คือค่าระดับความสนใจ (Interest Level - I) ที่มีต่อหมวดหมู่ (Category - c)

n_c คือจำนวนของการเช็คอินที่ปรากฏในหมวดหมู่ c

n_k คือจำนวนของการเช็คอินที่ปรากฏ ในทุก ๆ หมวดหมู่ k

2) โคลด์สตาร์ทระดับที่ 2 กรณีผู้ใช้มีการเช็คอินน้อยกว่า 10 จะถือว่ามีข้อมูลของผู้ใช้ไม่เพียงพอ เกิดปัญหาโคลด์สตาร์ทในระดับที่ 2 จะใช้ข้อมูลจากเพื่อนของผู้ใช้ในเฟสบุ๊คมาพิจารณาแทน ซึ่งเพื่อนมักจะมีความสนใจหลายอย่างเหมือนกันเมื่อมีพฤติกรรมที่คล้ายกัน (Ye, el al., 2011) เช่น เพื่อนที่ไปรับประทานอาหาร ชมภาพยนตร์ หรือ ท่องเที่ยวด้วยกัน ผู้วิจัยจึงพิจารณาถึงตัวแปรที่บอกถึงพฤติกรรมที่คล้ายกันบนเครือข่ายสังคมนั่นคือการทำกิจกรรมบางอย่างร่วมกันหรือการมีปฏิสัมพันธ์กัน ดังนั้นจึงสมมติฐานว่ากลุ่มเพื่อนที่มีการปฏิสัมพันธ์กับผู้ใช้ในระดับสูง จะมีความสนใจที่คล้ายกับผู้ใช้มากกว่ากลุ่มเพื่อนที่มีการปฏิสัมพันธ์กับผู้ใช้ในระดับต่ำ ผู้วิจัยจึงใช้ข้อมูลจากกลุ่มเพื่อนเพื่อแก้ไขปัญหาโคลด์สตาร์ทนี้ โดยเรียงลำดับเพื่อนตามระดับปฏิสัมพันธ์ที่มีกับผู้ใช้ (3.5) จากการประยุกต์ใช้อัลกอริทึมชื่อ ay-fb-friend-rank (gajus, 2012) ซึ่งอาศัยหลักการที่ชื่อว่า EdgeRank (Widman, 2014) โดย EdgeRank นั้นประกอบด้วย 3 ส่วน คือ 1) Affinity Score คือค่าที่คำนวณจากการปฏิสัมพันธ์ประเภทต่าง ๆ ที่มีระหว่างผู้ใช้ เช่น การกดไลค์ การแชร์ และการแสดงความคิดเห็นของเพื่อนบนโพสที่มีร่วมกัน การเช็คอินร่วมกัน การมีเพื่อนร่วมกัน เป็นต้น 2) Edge Weight คือค่าน้ำหนักที่ให้กับ Affinity Score แต่ละประเภท เช่น การแสดงความคิดเห็นจะมีค่าน้ำหนักสูงกว่าการกดไลค์ เป็นต้น และ 3) Time Decay คือค่าระยะเวลาที่โพสถูก สร้างขึ้น เช่น โพสที่ถูกสร้างล่าสุดจะมีค่าสูงกว่าโพสที่สร้างเมื่อสัปดาห์ที่แล้ว เป็นต้น โดยอัลกอริทึม ay-fb-friend-rank นั้นใช้เพียง 2 ส่วนประกอบ คือ 1) Affinity Score และ

2) Edge Weight เนื่องจากต้องการหลีกเลี่ยงการคำนวณ Time Decay ที่ต้องใช้เวลานาน

1. Affinity Score คือค่าการปฏิสัมพันธ์ที่มีระหว่างผู้ใช้และเพื่อน โดยในอัลกอริทึม ay-fb-friend-rank คำนวณจากการปฏิสัมพันธ์ 4 ประเภท ดังนี้

1.1 User's feed entries เช่น feed_like, feed_comment, feed_addressed เป็นต้น คือจะให้ค่าคะแนนกับเพื่อนที่มีการกดชื่นชอบ แสดงความเห็นลงบนหน้าเฟสบุ๊คของผู้ใช้ หรือติดแท็ก (tagging) ไปยังผู้ใช้

1.2 Photos เช่น photo_tagged_user_by_friend, photo_tagged_friend_by_user, photo_like, photo_comment เป็นต้น คือจะให้ค่าคะแนนกับเพื่อนที่มีการติดแท็กที่รูปภาพของผู้ใช้ หรือผู้ใช้ถูกติดแท็กไปยังรูปภาพของเพื่อน รวมถึงการที่เพื่อนหรือผู้ใช้กดชื่นชอบหรือแสดงความคิดเห็นบนรูปภาพเหล่านั้นด้วย

1.3 Friends เช่น friend_mutual เป็นต้น คือจะให้ค่าคะแนนไปยังเพื่อนที่มีเพื่อนร่วมกัน

1.4 Inbox เช่น inbox_in_conversation, inbox_chat เป็นต้น คือจะให้ค่าคะแนนไปยังเพื่อนที่มีการสนทนาระหว่างบุคคลกับผู้ใช้ (chat) หรือสนทนาเป็นกลุ่ม (conversation)

2. Edge Weights คือค่าน้ำหนักที่ให้กับแต่ละ Affinity Score โดยค่าน้ำหนักนี้แสดงถึงความสำคัญของปฏิสัมพันธ์แต่ละประเภทของ Affinity Score ซึ่งกำหนดให้ Affinity Score ที่มีค่า Edge Weights สูงกว่ามีความสำคัญมากกว่า อย่างไรก็ตามอัลกอริทึม ay-fb-friend-rank ให้ค่าน้ำหนักกับทุกประเภทเป็น 1 ทั้งหมด

เมื่อได้รายชื่อเพื่อนสนิทและค่าปฏิสัมพันธ์ระหว่างผู้ใช้กับเพื่อนจากการคำนวณโดยอัลกอริทึมข้างต้น ซึ่งจากผลการทดลองพบว่าการเลือกใช้เพื่อน 5 อันดับแรก (กลุ่มเพื่อน) มีความเหมาะสมที่สุด โดยหาก 5 อันดับแรกมีข้อมูลการเช็คอินไม่เพียงพอ ก็จะเลือกลำดับถัดไปที่มีข้อมูลเพียงพอจนครบ 5 คน ซึ่งหากสุดท้ายแล้ว 3 ใน 5 ยังมีข้อมูลเช็คอินเพียงพอ จะใช้ข้อมูลเช็คอินจากเพื่อนคำนวณค่าระดับความสนใจ โดยวิธีการเช่นเดียวกับการคำนวณไปยังผู้ใช้ในสมการที่ 2 สุดท้ายใช้สมการที่ 2 คำนวณหาค่าระดับความสนใจของผู้ใช้จากค่าระดับความสนใจของเพื่อน

$$I(c)=\frac{\sum_{i=1}^{n} F_i \; I_i(c)}{\sum_{i=1}^{n} F_i}, \; n=5 \qquad (2)$$

$I(c)$ คือค่าระดับความสนใจ (Interest Level - I) ที่มีต่อหมวดหมู่ (Category - c)

F_i คือค่าปฏิสัมพันธ์ระหว่างผู้ใช้และเพื่อนคนที่ i

$I_i(c)$ คือค่าระดับความสนใจ (Interest Level - I) ที่มีต่อหมวดหมู่ (Category - c) ของเพื่อนคนที่ i

3) ใช้ข้อมูลของกลุ่มตัวอย่าง โดยจะใช้ในกรณีที่ไม่สามารถแก้ไขโคล์ดสตาร์ทระดับที่ 2 ได้ จะใช้ค่าเฉลี่ยของค่าระดับความสนใจในแต่ละหมวดหมู่ที่คำนวณจากข้อมูลการเช็คอินของผู้ใช้กลุ่มตัวอย่าง โดยตัวอย่างจะเลือกจากผู้ใช้ที่มีเช็คอินที่ผ่านการคลีนซิงมากกว่า 10 เช็คอิน

วิธีคำนวณค่าระดับความสนใจมีการทำงานดังอัลกอริทึมที่ ในตารางที่ 7 และค่าระดับความสนใจที่คำนวณได้ จะส่งไปยังกระบวนการถัดไป เพื่อใช้ในการแนะนำสถานที่ท่องเที่ยว

ตาราง 7 วิธีคำนวณค่าระดับความสนใจของผู้ใช้จากการใช้ข้อมูลเช็คอิน

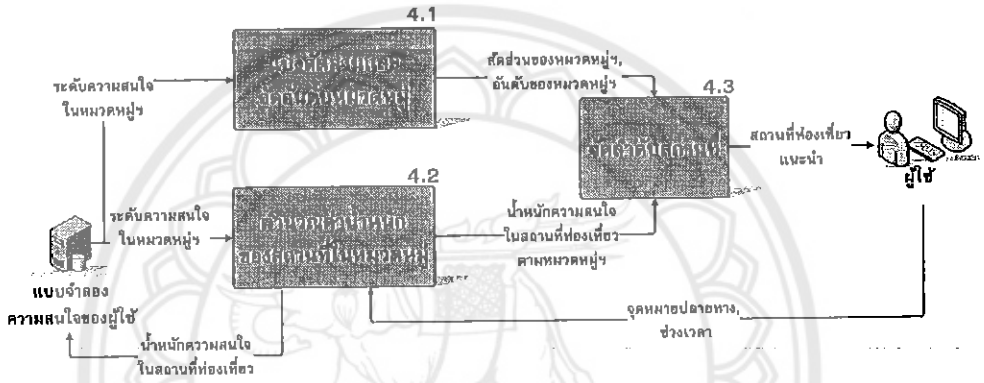
อัลกอริทึมที่ 1: วิธีคำนวณค่าระดับความสนใจของผู้ใช้จากการใช้ข้อมูลเช็คอิน

```
1 :  ดึงข้อมูลสถานที่เช็คอินของผู้ใช้จากเฟสบุ๊ค
2 :  ลบสถานที่เช็คอินที่ซ้ำซ้อนออก
3 :  ลบสถานที่เช็คอินที่ไม่ใช่จุดประสงค์เพื่อการท่องเที่ยวออก
4 :  แบ่งหมวดหมู่ของสถานที่เช็คอิน
5 :    if สถานที่เช็คอินที่แบ่งหมวดหมู่แล้ว รวมมีทั้งหมด > 10
6 :      คำนวณค่าระดับความสนใจ I(c) ในสมการที่ 2
7 :    else
8 :      เรียงลำดับเพื่อนในเฟสบุ๊คตามระดับปฏิสัมพันธ์ที่มีกับผู้ใช้
9 :      ดึงข้อมูลสถานที่เช็คอินของเพื่อน 5 คนจากเฟสบุ๊ค
10 :     if สถานที่เช็คอินที่แบ่งหมวดหมู่แล้วของเพื่อนอย่างน้อย 3 คน
           โดยแต่ละคนรวมมีทั้งหมด > 10
11 :       คำนวณค่าระดับความสนใจ I(c) ในสมการที่ 3
```

อัลกอริทึมที่ 1: วิธีคำนวณค่าระดับความสนใจของผู้ใช้จากการใช้ข้อมูลเช็คอิน

```
12 :        else
13 :            ใช้ค่าเฉลี่ยของค่าระดับความสนใจ I(c) ของกลุ่มตัวอย่าง
14 :    ส่งค่า I(c) ไปยังขั้นตอนต่อไป
```

3.5 การแนะนำข้อมูลการท่องเที่ยว (Recommendation)



ภาพ 6 การแนะนำข้อมูลการท่องเที่ยว

เมื่อได้ข้อมูลความสนใจที่วิเคราะห์จากเครือข่ายสังคมแล้ว กระบวนการต่อไปเป็นการแนะนำข้อมูลการท่องเที่ยวไปยังผู้ใช้ ดังภาพที่ 6 โดยใช้วิธีการแนะนำข้อมูลแบบผสมผสาน (Hybrid) ทั้งสองวิธีคือ Content-based และ Collaborative ดังนี้

1. โคลด์สตาร์ทระดับที่ 1 กรณีที่ระบบไม่มีข้อมูลความสนใจของผู้ใช้ที่เริ่มใช้งานระบบ ใช้หลักการ User-based Content Filtering ซึ่งใช้การวิเคราะห์ความสนใจที่ได้จากการเช็คอินในอดีตและแนะนำข้อมูลการท่องเที่ยวไปยังผู้ใช้

2. โคลด์สตาร์ทระดับที่ 2 กรณีที่ไม่สามารถใช้ข้อมูลการเช็คอินของผู้ใช้ในโคลด์สตาร์ทในระดับที่ 1 ได้ โดยใช้หลักการ Friend-based Collaborative Filtering ซึ่งจะใช้ข้อมูลการเช็คอินจากเพื่อนที่มีพฤติกรรมหรือความสนใจคล้ายกันในการวิเคราะห์ความสนใจและแนะนำข้อมูลการท่องเที่ยวไปยังผู้ใช้

กระบวนการในการแนะนำข้อมูล แบ่งออกเป็น 3 ขั้นตอน

QA
76
.ท7158
อ157ร
2560
1020159
สำนักหอสมุด
พ.ศ. 2562

3.5.1 การแบ่งสัดส่วนและจัดอันดับหมวดหมู่ โดยแบ่งสัดส่วนในการแนะนำตามค่าระดับความสนใจที่คำนวณได้ในกระบวนการก่อนหน้า หมวดหมู่ที่ผู้ใช้มีความสนใจสูงก็จะถูกแนะนำในสัดส่วนที่มากและถูกจัดอันดับไว้สูงกว่าหมวดหมู่ที่ผู้ใช้มีความสนใจน้อยกว่าโดยการคำนวณสัดส่วนแสดงดังสมการที่ 3

$$R(c)=\frac{I(c)}{\sum_{k=1}^{6} I(k)}\times N,\ 1\le c\le 6 \qquad (3)$$

$R(c)$ คือค่าสัดส่วนการแนะนำ (Ratio - R) ที่มีต่อหมวดหมู่ (Category - c)

$I(c)$ คือค่าระดับความสนใจ (Interest Level - I) ที่มีต่อหมวดหมู่ (Category - c)

N คือจำนวนของสถานที่ที่ต้องการแสดงไปยังผู้ใช้

3.5.2 คำนวณค่าน้ำหนักของสถานที่ในแต่ละหมวดหมู่ เป็นการคำนวณค่าน้ำหนักของสถานที่ที่อยู่ในแต่ละหมวดหมู่จากฐานข้อมูลสถานที่ท่องเที่ยว โดยเลือกพิจารณาเฉพาะปลายทางที่ผู้ใช้ระบุ เช่น ปลายทางคือจังหวัดพิษณุโลก ก็จะนำสถานที่ที่อยู่ในจังหวัดพิษณุโลกมาคำนวณเท่านั้น ซึ่งเกณฑ์การคำนวณแบ่งเป็น 3 ส่วน คือ 1) ความนิยมของสถานที่ 2) เป็นสถานที่ที่เพื่อน ๆ เคยไป และ 3) เป็นช่วงเวลาที่นิยมไป ดังสมการที่ 4

$$W(p)=\alpha P(p)+\beta F(p)+\gamma T(p) \qquad (4)$$

$W(p)$ คือค่าน้ำหนักของสถานที่ (Place Weight - W) ที่มีต่อสถานที่ (Place - p)

$P(p)$ คือค่าความนิยม (Popularity - P) ของสถานที่ (Place - p)

$F(p)$ คือค่าคะแนน (Friend - F) ของสถานที่ (Place - p) ที่เพื่อนสนิทในเฟสบุ๊คของผู้ใช้เคยไป

$T(p)$ คือค่าช่วงเวลาที่เหมาะสม(Time - T) ที่มีต่อสถานที่ (Place - p) โดยมีค่า =1 เมื่อมีสถานที่ (Place - p) เหมาะสมที่จะไปในช่วงเวลาที่ผู้ใช้กำหนด และมีค่า = 0 เมื่อไม่เหมาะสม

α คือค่าน้ำหนักของค่าความนิยม

β คือค่าน้ำหนักของการคำนวณสถานที่จากเพื่อน

γ คือค่าน้ำหนักของค่าเวลา

โดย $0<\alpha+\beta+\gamma<1$ และค่าที่เหมาะสมของทั้ง α, β และ γ แสดงในผลการทดลองที่ 4

ทั้งนี้ค่าความนิยม $P(p)$ คิดจากจำนวนการเช็คอินและการกดไลค์จากผู้ใช้ทั้งหมดบนเฟสบุ๊ค ซึ่งคำนวณดังสมการที่ 5 โดยหากการเช็คอินและกดไลค์มีจำนวนเกินกว่า 10,000 จะให้ค่า $\frac{n_{ch,p}}{\max(n_{ch,p})}$ และ $\frac{n_{li,p}}{\max(n_{li,p})}$ เท่ากับ 1

$$P(p)=\alpha_{ch}\frac{n_{ch,p}}{\max(n_{ch,p})}+\alpha_{li}\frac{n_{li,p}}{\max(n_{li,p})} \quad (5)$$

$P(p)$ คือค่าความนิยม (Popularity - P) ของสถานที่ (Place - p)

α_{ch} คือค่าน้ำหนักของการเช็คอิน (Check-in - ch) ที่มีต่อสถานที่ (Place - p)

α_{li} คือค่าน้ำหนักการกดไลค์ (Like - li) ที่มีต่อสถานที่ (Place - p)

$n_{ch,p}$ คือจำนวนของการเช็คอิน (Check-in - ch) ที่มีต่อสถานที่ (Place - p)

$n_{li,p}$ คือจำนวนการกดไลค์ (Like - li) ที่มีต่อสถานที่ (Place - p)

$n_{ch,p}$ คือจำนวนของการเช็คอิน (Check-in - ch) ที่มีต่อทุก ๆ สถานที่ k

$n_{li,p}$ คือจำนวนของการกดไลค์ (Like - li) ที่มีต่อทุก ๆ สถานที่ k

โดยกำหนดให้ α_{ch} และ α_{li} เท่ากับ 0.5

และค่าคะแนนของสถานที่ที่เพื่อนสนิทในเฟสบุ๊คของผู้ใช้เคยไป $F(p)$ คำนวณดังสมการที่ 6 โดยจะไม่ใช่ค่านี้คำนวณร่วมในสมการที่ 5 ในกรณีที่ใช้การคำนวณค่า $I(c)$ จากกลุ่มตัวอย่าง (3.8)

$$F(p)=\frac{\sum_{i=1}^{n}F_iC_i}{\sum_{i=1}^{n}F_i},n=5 \quad (6)$$

$F(p)$ คือ ค่าคะแนนของสถานที่ (Place - p) ที่เพื่อนสนิทในเฟสบุ๊คของผู้ใช้เคยไป

F_i คือ ค่าปฏิสัมพันธ์ระหว่างผู้ใช้และเพื่อนคนที่ i

C_i มีค่า = 1 เมื่อมีสถานที่ (Place - p) ปรากฏในการเช็คอินของเพื่อนคนที่ i และมีค่า = 0 เมื่อไม่ปรากฏ

3.5.3 จัดลำดับสถานที่ เป็นการจัดลำดับ (Rank) สถานที่ที่จะแนะนำไปยังผู้ใช้ โดยใช้ค่าสัดส่วนของหมวดหมู่ $R(c)$ และค่าน้ำหนัก $W(p)$ ที่คำนวณไปยังสถานที่ ดังสมการที่ 7 และกำหนดให้ ถ้าค่า $Rank(p_1)$ ค่าสูงกว่า $Rank(p_2)$ แล้ว p_1 จะมีลำดับสูงกว่า p_2

$$Rank(p) = \frac{R(c) \times W(p)}{N} \quad (7)$$

$Rank(p)$ คือค่าลำดับของสถานที่ (Place - p)

$R(c)$ คือค่าสัดส่วนการแนะนำ (Ratio - R) ที่มีต่อหมวดหมู่ (Category - c) ของสถานที่ (Place - p)

$W(p)$ คือค่าน้ำหนักของสถานที่ (Place Weight -W) ที่มีต่อสถานที่ (Place - p)

N คือจำนวนของสถานที่ที่ต้องการแสดงไปยังผู้ใช้

กระบวนการแนะนำข้อมูลการท่องเที่ยวใช้อัลกอริทึมที่ 2 ในตารางที่ 8 ดังนี้

ตาราง 8 การแนะนำสถานที่ท่องเที่ยว

อัลกอริทึมที่ 2 : การแนะนำสถานที่ท่องเที่ยว	
1 :	คำนวณสัดส่วนการแนะนำ $R(c)$ ในสมการที่ 4
2 :	คำนวณค่าน้ำหนักของสถานที่ $W(p)$ ในสมการที่ 5
3 :	คำนวณค่าลำดับของสถานที่ $Rank(p)$ ในสมการที่ 8
4 :	แนะนำสถานที่ท่องเที่ยวจากลำดับของสถานที่ $Rank(p)$

บทที่ 4
ผลการวิจัย

ในบทนี้จะอธิบายถึงผลการดำเนินการวิจัยโดยประกอบด้วย 3 ส่วน คือ ผลการวิเคราะห์ความต้องการของผู้ใช้ ผลการออกแบบและพัฒนาระบบ และผลการทดสอบระบบ ซึ่งจะอธิบายดังต่อไปนี้

4.1 ผลการวิเคราะห์ความต้องการของผู้ใช้

1. การเก็บรวบรวมความต้องการ

การเก็บรวบรวมความต้องการของผู้ใช้ในงานวิจัยนี้กระทำโดยให้ผู้ใช้ตอบแบบสอบถามความต้องการสำหรับระบบให้บริการข้อมูลการเดินทางเฉพาะบุคคลสำหรับการท่องเที่ยวไทยแบบออนไลน์ โดยมีจำนวนผู้ตอบแบบสอบถามทั้งสิ้นเป็นจำนวน 450 คน แบ่งเป็นเพศชาย 191 คน (42.2%) และเพศหญิง 259 คน (57.8%) โดยผู้ตอบแบบสอบถามส่วนใหญ่มีช่วงอายุระหว่าง 31-45 ปีคิดเป็น 42.7% และอายุระหว่าง 18-30 ปี คิดเป็น (38.7%) ซึ่งเป็นช่วงอายุที่ใช้คอมพิวเตอร์และอินเทอร์เน็ตในชีวิตประจำวันทุกวัน สำหรับผู้ตอบแบบสอบถามที่มีอายุเกิน 60 ปี มีจำนวนเพียง 6.7% เท่านั้นเนื่องจากคนในช่วงอายุดังกล่าวอาจจะไม่ใช้อินเทอร์เน็ตในชีวิตประจำวันมากนัก

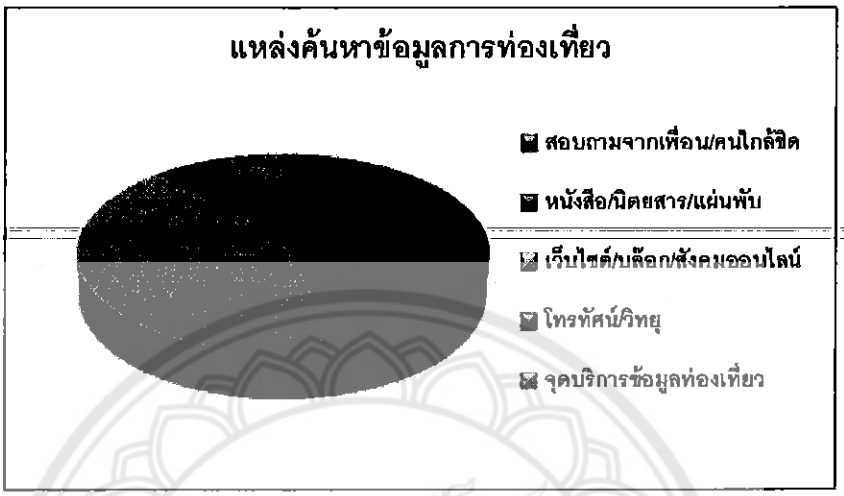
2. ผลการวิเคราะห์ความต้องการของผู้ใช้

จากการเก็บรวบรวมข้อมูลจากแบบสอบถามสามารถสรุปได้ดังต่อไปนี้

2.1 แหล่งข้อมูลการท่องเที่ยวที่ใช้ในการค้นหาข้อมูล

จากข้อมูลที่รวบรวมได้จะพบว่าแหล่งค้นหาข้อมูลที่สำคัญของผู้ตอบแบบสอบถามคือแหล่งข้อมูลบนเครือข่ายอินเทอร์เน็ต (ภาพที่ 7) เช่น เว็บไซต์ต่าง ๆ บล็อก (Blog) และสังคมออนไลน์ ทั้งนี้เนื่องจากมีความสะดวก รวดเร็วและให้ข้อมูลที่หลากหลายต่อนักท่องเที่ยว นอกจากนี้จากการสำรวจยังพบอีกว่า นักท่องเที่ยวมากกว่าร้อยละ 59% จะใช้เครือข่ายอินเทอร์เน็ตในการจองโรงแรม ซื้อตั๋วเครื่องบิน หรือธุรกรรมการท่องเที่ยวอื่น ๆ ดังนั้นจึงสรุปได้ว่าเครือข่ายอินเทอร์เน็ตเป็นแหล่งข้อมูลที่สำคัญในการให้ข้อมูลข่าวสารด้านการท่องเที่ยวแก่นักเดินทางทั้งในและต่างประเทศ ข้อมูลที่น่าสนใจอีกประการหนึ่งคือ นอกจากแหล่งข้อมูลบนเครือข่ายอินเทอร์เน็ตแล้ว นักท่องเที่ยวส่วนมากมักจะถามข้อมูลการท่องเที่ยวจากเพื่อน ๆ ที่เคยไปสถานที่ท่องเที่ยวต่าง ๆ มาก่อน (29.5%) ดังนั้นจะเห็นว่าข้อมูลจากเพื่อนที่เคยไปเที่ยวตามแหล่ง

ท่องเที่ยวต่าง ๆ มาก่อนเป็นส่วนสำคัญในการตัดสินใจเดินทางไปสถานที่ต่าง ๆ ของตัวนักท่องเที่ยวเองเช่นกัน

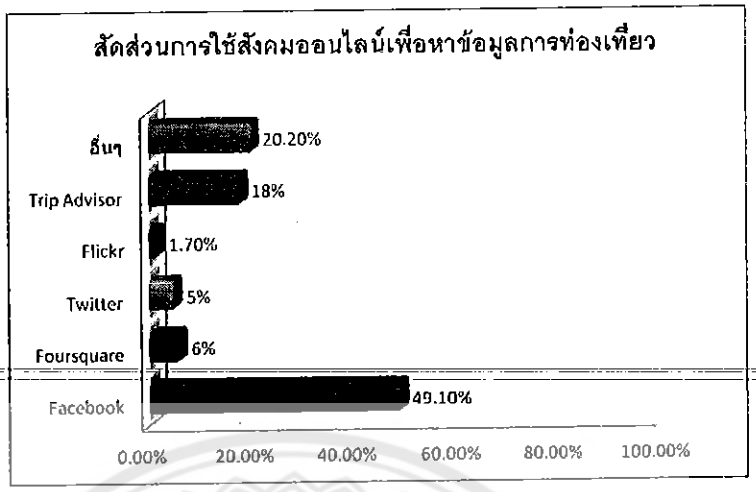


ภาพ 7 แหล่งค้นหาข้อมูลสถานที่ท่องเที่ยวต่าง ๆ

2.2 การใช้สังคมออนไลน์เพื่อเป็นแหล่งข้อมูลการท่องเที่ยว

จากการสอบถามกลุ่มตัวอย่างเกี่ยวกับการใช้สื่อสังคมออนไลน์ พบว่าผู้ตอบแบบสอบถามส่วนใหญ่ (84.7%) ใช้สังคมออนไลน์เพื่อหาข้อมูลท่องเที่ยว และในจำนวนสื่อสังคมออนไลน์ต่าง ๆ พบว่า เฟสบุ๊ค เป็นสังคมออนไลน์อันดับหนึ่งที่มีผู้ใช้มากที่สุด ประมาณ 49% (ภาพที่ 8) รองลงมาคือ Trip Advisor (18%) ดังนั้นจึงสรุปได้ว่าเฟสบุ๊คจะเป็นแหล่งข้อมูลที่มีประโยชน์สำหรับการให้ข้อมูลการท่องเที่ยวกับผู้ใช้ได้

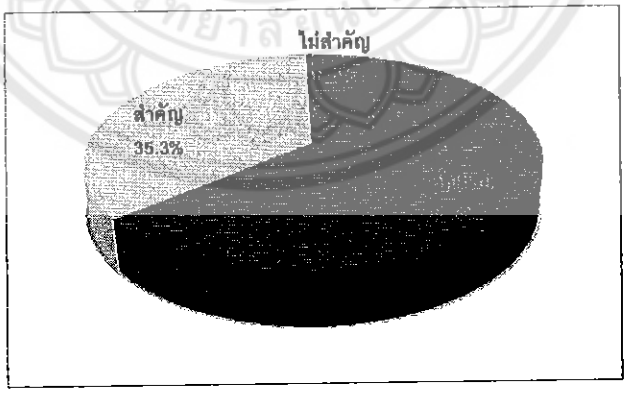
นอกจากนี้ผู้ตอบแบบสอบถามส่วนใหญ่ (92.7%) ยังสนใจสถานที่ท่องเที่ยวที่เพื่อนในกลุ่มเคยไปมาและสนใจที่จะไปสถานที่ดังกล่าว ดังนั้นหากสามารถตรวจสอบได้ว่าเพื่อน ๆ ของนักท่องเที่ยวเคยไปที่ไหนบ้างและนำมาแนะนำแก่นักท่องเที่ยว จะทำให้ระบบสามารถแนะนำข้อมูลได้ตรงกับความต้องการของนักท่องเที่ยวมากยิ่งขึ้น ช่องทางหนึ่งที่สามารถตรวจสอบได้คือผ่านช่องทางสื่อสังคมออนไลน์นั่นเอง และสถิตินี้มีความสอดคล้องกับหัวข้อ 2.1 นั่นคือข้อมูลที่ได้จากเพื่อน ๆ เป็นข้อมูลสำคัญที่นักท่องเที่ยวใช้เพื่อตัดสินใจว่าจะไปสถานที่ต่าง ๆ หรือไม่



ภาพ 8 สัดส่วนการใช้สังคมออนไลน์เพื่อหาข้อมูลการท่องเที่ยว

2.3 การแนะนำข้อมูลสถานที่ท่องเที่ยวเฉพาะบุคคล

นอกจากนี้ผู้ตอบแบบสอบถามยังมีความเห็นว่า ระบบที่ใช้ข้อมูลการท่องเที่ยวควรสามารถให้ข้อมูลที่ตรงกับความต้องการของนักท่องเที่ยวมากที่สุด โดยเห็นว่าระบบดังกล่าวมีความสำคัญมากถึง 64.2% และมีเพียง 0.4% ที่คิดว่าระบบดังกล่าวไม่สำคัญ (ภาพที่ 9) ดังนั้นจึงสรุปได้ว่า ผู้ใช้เห็นว่าการพัฒนาระบบให้ข้อมูลการท่องเที่ยวเฉพาะบุคคลมีความสำคัญและเป็นประโยชน์ต่อนักท่องเที่ยวอย่างยิ่ง



ภาพ 9 ความสำคัญในการพัฒนาระบบการให้ข้อมูลการท่องเที่ยวเฉพาะบุคคล

4.2 ผลการออกแบบและพัฒนาระบบ

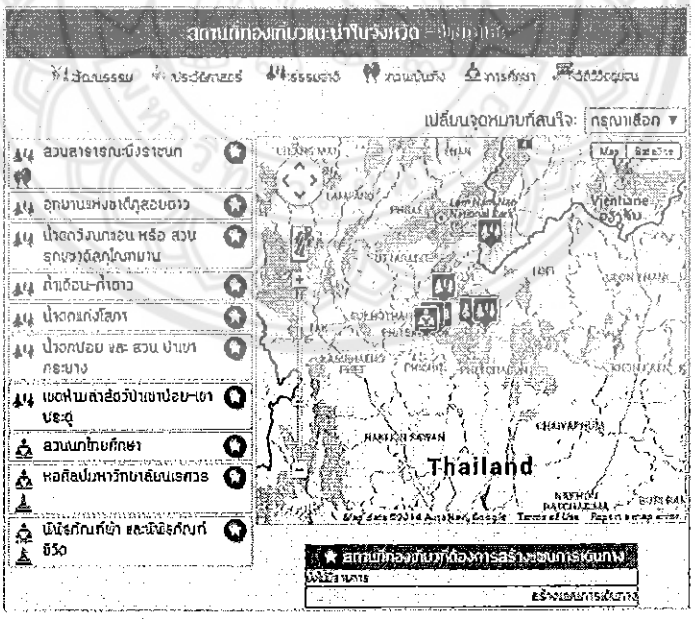
การแนะนำสถานที่ท่องเที่ยวจากระบบ จะแบ่งการทำงานออกเป็น 2 แบบ

1. การแนะนำสถานที่จากระบบสำหรับผู้ใช้แบบผู้มาเยือน (Guest)

ผู้ใช้สามารถใช้ระบบแนะนำสถานที่ได้ โดยไม่ต้องล็อกอิน โดยระบบจะเก็บข้อมูลการแนะนำไว้จนกระทั่งผู้ใช้ปิดเว็บบราวเซอร์ ยกตัวอย่างเช่น เริ่มต้นคือเลือกจุดหมายปลายทางและช่วงเวลา ดังภาพที่ 10

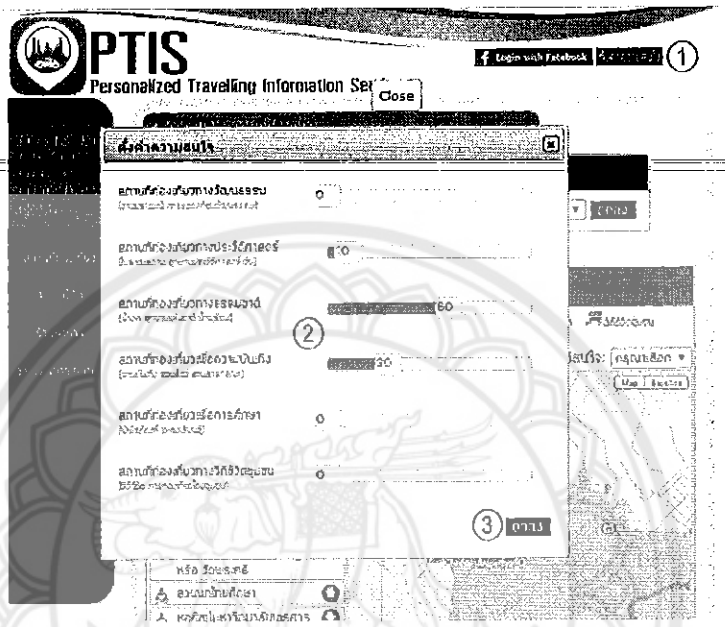
ภาพ 10 การคลิกเพื่อเลือกจุดหมายปลายทางและช่วงเวลา

จากนั้นระบบจะแนะนำสถานที่โดยวิเคราะห์ความสนใจจากค่าเฉลี่ยของกลุ่มตัวอย่างและ จุดหมายปลายทางและช่วงเวลาที่ระบุจากขั้นตอนก่อนหน้า ดังภาพที่ 11



ภาพ 11 หน้าจอแนะนำสถานที่ท่องเที่ยว

โดยผู้ใช้สามารถปรับค่าความสนใจ เพื่อให้ระบบวิเคราะห์และแนะนำใหม่ได้ โดยคลิกที่ "ตั้งค่าความสนใจ" จากนั้นทำการปรับเปลี่ยนความสนใจในสไลด์บาร์ และคลิก "ตกลง" ดังภาพที่ 12



ภาพ 12 หน้าจอการปรับค่าความสนใจ

ผู้ใช้สามารถเลือกสถานที่เพื่อนำไปสร้างแผนการเดินทางได้ ดังภาพที่ 13 โดยคลิกที่ปุ่มรูปดาว สถานที่ที่ต้องการสร้างแผนการเดินทางจะปรากฏในตารางด้านล่าง และสามารถลบออกได้โดยคลิกปุ่มกากบาทในสถานที่ที่ต้องการนำออก นอกจากนี้สามารถ "เปลี่ยนจุดหมายที่สนใจ" ซึ่งจุดหมายที่สนใจจะแสดงจังหวัดที่เดินทางผ่านจากจังหวัดต้นทางและปลายทางที่ผู้ใช้ระบุ โดยผู้ใช้สามารถเลือกจุดหมายที่สนใจใหม่ได้ในรายการตัวเลือกที่แสดง

ภาพ 13 เลือกสถานที่เพื่อสร้างแผนการเดินทาง

2. การแนะนำสถานที่จากระบบสำหรับผู้ใช้แบบสมาชิก (Member)

ผู้ใช้ที่มีบัญชีเฟสบุ๊คและล็อกอินเข้าสู่ระบบ ระบบจะให้คำแนะนำสถานที่โดยสมบูรณ์ โดยระบบจะวิเคราะห์ความสนใจของผู้ใช้จากเครือข่ายสังคมออนไลน์ จากนั้นจะใช้ข้อมูลตอบกลับที่ผู้ใช้ทำกับระบบ เช่น การคลิกดูรายละเอียด การเลือกเป็นสถานที่ที่ต้องการสร้างแผนการเดินทาง เป็นต้น ปรับปรุงคำแนะนำและความสนใจของผู้ใช้ใหม่เพื่อให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น ข้อมูลความสนใจต่าง ๆ ระบบจะเก็บไปยังแบบจำลองผู้ใช้ เมื่อผู้ใช้กลับมาล็อกอินเข้าสู่ระบบอีกครั้งระบบจะดึงข้อมูลจากแบบจำลองผู้ใช้มาแนะนำสถานที่ได้ทันที

ในกรณีที่ผู้ใช้เริ่มต้นใช้งานระบบครั้งแรก ระบบจะให้ผู้ใช้ประมวลผลข้อมูลจากเฟสบุ๊คเพื่อวิเคราะห์ระดับความสนใจในภาพที่ 14 เมื่อประมวลผลเรียบร้อยแล้ว ทำการเลือกจุดหมายปลายทางและช่วงเวลา ที่ต้องการให้แนะนำ จากนั้นระบบจะแนะนำสถานที่โดยวิเคราะห์ความสนใจจากระดับความสนใจที่วิเคราะห์มาจากเฟสบุ๊ค ดังภาพที่ 15

ภาพ 14 ปุ่มเริ่มประมวลผลข้อมูลจากเฟสบุ๊ค

ภาพ 15 ระบบแนะนำสถานที่ท่องเที่ยวไปยังผู้ใช้

ทั้งนี้ผู้ใช้สามารถเปลี่ยนแปลงค่าระดับความสนใจเหล่านี้ใหม่ได้ โดยคลิกที่ปุ่ม "ตั้งค่าความสนใจ" เช่นกัน ดังภาพที่ 12 ซึ่งการเปลี่ยนแปลงค่าก็จะทำให้แบบจำลองความสนใจเปลี่ยนไป

จากภาพที่ 15 ผู้ใช้สามารถเลือกสถานที่เพื่อสร้างแผนการเดินทางโดยคลิกปุ่มรูปดาว สถานที่ที่ต้องการสร้างแผนการเดินทางจะปรากฏในตารางด้านล่าง และสามารถลบออกได้โดยคลิกปุ่มกากบาทในสถานที่ที่ต้องการนำออก และผู้ใช้สามารถกดปุ่มรูปกากบาท ซึ่งหมายถึงไม่สนใจในสถานที่นี้ ต้องการให้ระบบแนะนำสถานที่อื่น ๆ แทน โดยพฤติกรรมที่ผู้ใช้มีต่อระบบเหล่านี้

เป็นข้อมูลตอบกลับที่จะนำไปคำนวณเพื่อปรับปรุงค่าความสนใจในหมวดหมู่และค่าความสนใจในสถานที่ของผู้ใช้ โดยผู้ใช้สามารถกด "แนะนำสถานที่ใหม่" เพื่อให้ระบบคำนวณและสร้างคำแนะนำสถานที่ใหม่ให้ได้ตรงกับความต้องการของผู้ใช้ยิ่งขึ้น

4.3 ผลการประเมินระบบ

การประเมินระบบออกเป็น 2 ส่วน คือ 1) การประเมินประสิทธิภาพของระบบโดยกลุ่มตัวอย่าง และ 2) การประเมินความพึงพอใจของผู้ใช้งานผ่านเว็บแอปพลิเคชัน PTIS ซึ่งมีรายละเอียดดังนี้

1. การประเมินประสิทธิภาพของระบบโดยกลุ่มตัวอย่าง

การประเมินประสิทธิภาพของระบบ ได้เลือกกลุ่มตัวอย่างจำนวน 42 ตัวอย่าง เป็นนักท่องเที่ยวที่มีความสนใจหรือวางแผนจะท่องเที่ยวในจังหวัดต้นแบบ 5 จังหวัด ในอนาคต มีการใช้อินเทอร์เน็ตเป็นแหล่งค้นหาข้อมูลเพื่อวางแผนการท่องเที่ยว โดยมีบัญชีผู้ใช้เฟสบุ๊คที่มีการใช้งานอย่างน้อย 1 ครั้งต่อสัปดาห์และไม่มีภูมิลำเนาอยู่ในจังหวัดต้นแบบ ลักษณะทั่วไปของกลุ่มตัวอย่างเป็นดังนี้

1.1 ข้อมูลทั่วไปของกลุ่มตัวอย่าง

จากกลุ่มตัวอย่างจำนวน 42 ตัวอย่าง พบว่าส่วนมากเป็นเพศหญิง โดยมีจำนวน 26 คน คิดเป็นร้อยละ 61.90 เป็นเพศชาย จำนวน 16 คน คิดเป็นร้อยละ 38.10 ดังแสดงในตารางที่ 9

ตาราง 9 ลักษณะของกลุ่มตัวอย่าง

เพศ	จำนวน	ร้อยละ	แปลผล
ชาย	16	38.10	น้อย
หญิง	26	61.90	มาก
รวม	42	100.00	หญิงมากกว่าชาย

จากกลุ่มตัวอย่างจำนวน 42 ตัวอย่าง พบว่าช่วงอายุการศึกษาของกลุ่มตัวอย่างส่วนมาก คือ มีช่วงอายุ 18 - 30 ปี จำนวน 32 คน คิดเป็นร้อยละ 76.19 รองลงมาช่วงอายุ 31 - 45 ปี จำนวน 9 คน คิดเป็นร้อยละ 21.43 และช่วงอายุ 46 - 60 ปี จำนวน 1 คน คิดเป็นร้อยละ 2.38 ตามลำดับ ดังแสดงในตาราง 10

ตาราง 10 ข้อมูลช่วงอายุของกลุ่มตัวอย่าง

ช่วงอายุ	จำนวน	ร้อยละ	แปลผล
น้อยกว่า 18 ปี	0	0.00	ไม่มี
18 - 30 ปี	32	76.19	มาก
31 - 45 ปี	9	21.43	น้อย
46 - 60 ปี	1	2.38	น้อยมาก
มากกว่า 60 ปี	0	0.00	ไม่มี
รวม	42	100.00	มีช่วงอายุ 18 - 30 ปี เป็นส่วนมาก

จากกลุ่มตัวอย่างจำนวน 42 ตัวอย่าง ซึ่งไม่มีภูมิลำเนาอยู่ในจังหวัดต้นแบบที่ระบบแนะนำสถานที่ท่องเที่ยว ได้แก่ พิษณุโลก ตาก เพชรบูรณ์ อุตรดิตถ์ และสุโขทัย พบว่าส่วนมากไม่เคยมาท่องเที่ยวในจังหวัดต้นแบบ มีจำนวน 30 คน คิดเป็นร้อยละ 71.43 เคยมาท่องเที่ยวในจังหวัดต้นแบบ จำนวน 12 คน คิดเป็นร้อยละ 28.57 ดังแสดงในตาราง 11

ตาราง 11 ข้อมูลการมาท่องเที่ยวในจังหวัดต้นแบบ

การท่องเที่ยว	จำนวน	ร้อยละ	แปลผล
ไม่เคยมาท่องเที่ยว	30	71.43	มาก
เคยมาท่องเที่ยว	12	28.57	น้อย
รวม	42	100.00	ไม่เคยมาท่องเที่ยวจังหวัดต้นแบบเป็นส่วนมาก

จากกลุ่มตัวอย่างจำนวน 42 ตัวอย่าง พบว่าระดับการศึกษาของกลุ่มตัวอย่างส่วนมาก คือปริญญาตรี โดยมีระดับการศึกษาปริญญาตรี จำนวน 27 คน คิดเป็นร้อยละ 64.29 รองลงมาคือระดับการศึกษาปริญญาโท มีจำนวน 11 คน คิดเป็นร้อยละ 26.19 และมีระดับการศึกษาปริญญาเอกและระดับมัธยมศึกษา / อาชีวศึกษา จำนวน 2 คน คิดเป็นร้อยละ 4.76 ดังแสดงในตาราง 12

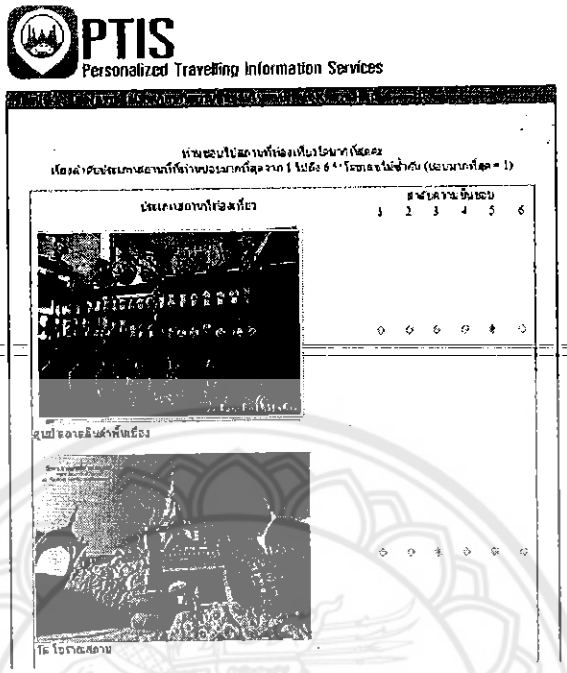
ตาราง 12 ข้อมูลระดับการศึกษาของกลุ่มตัวอย่าง

ระดับการศึกษา	จำนวน	ร้อยละ	แปลผล
มัธยมศึกษา / อาชีวศึกษา	2	4.76	ไม่มี
ประกาศนียบัตร / อนุปริญญา / ปวส.	0	0.00	ไม่มี
ปริญญาตรี	27	64.29	มาก
ปริญญาโท	11	26.19	มาก
ปริญญาเอก	2	4.76	น้อย
รวม	42	100.00	ระดับปริญญาตรีส่วนมาก

1.2 การประเมินประสิทธิภาพในการแก้ไขปัญหาโคลด์สตาร์ท (Cold Start Problem Evaluation)

การทดสอบนี้เป็นการวัดประสิทธิภาพในการแก้ไขปัญหาโคลด์สตาร์ทสองระดับ โดยวัดประสิทธิภาพจากค่าความถูกต้อง (Accuracy) ของลำดับความสนใจในหมวดหมู่การท่องเที่ยวที่วิเคราะห์ได้จากระบบ เปรียบเทียบกับลำดับความสนใจจริงที่ได้จากผู้ใช้จากแบบสอบถามลำดับความสนใจในภาพที่ 16 รวมถึงพิจารณาปริมาณของข้อมูลเช็คอินที่ส่งผลต่อความถูกต้องของการลำดับความสนใจ

ในการทดลองใช้กลุ่มตัวอย่างจำนวน 42 บัญชีผู้ใช้ ที่มีข้อมูลการเช็คอินที่ตรวจสอบความถูกต้องแล้วมากกว่า 30 เช็คอิน จากนั้นจำลองการเกิดปัญหาโคลด์สตาร์ททั้งสองระดับโดย 1) เริ่มต้นระบบไม่มีข้อมูลความสนใจ เกิดโคลด์สตาร์ทระดับ 1 และ 2) เริ่มต้นระบบไม่มีข้อมูลความสนใจของผู้ใช้และผู้ใช้มีเช็คอินไม่เพียงพอ เกิดโคลด์สตาร์ทระดับ 2 ในกรณีนี้ผู้วิจัยใช้การลบข้อมูลการเช็คอินของผู้ใช้ในเครือข่ายสังคมออกเพื่อให้มีข้อมูลไม่เพียงพอ ทำให้ระบบต้องใช้ข้อมูลของเพื่อนผู้ใช้ในการวิเคราะห์เพื่อสร้างลำดับความสนใจแทน จากนั้นเปรียบเทียบกับลำดับความสนใจแท้จริงที่ได้จากการสอบถามผู้ใช้ โดยตารางที่ 13 เป็นการเปรียบเทียบลำดับความสนใจของผู้ใช้ 1 คน กรณีโคล์ดสตาร์ทระดับที่ 1 และตารางที่ 14 เป็นการเปรียบเทียบลำดับความสนใจของผู้ใช้ 1 คน กรณีโคล์ดสตาร์ทระดับที่ 2



ภาพ 16 แบบสอบถามลำดับความสนใจ

ตาราง 13 เปรียบเทียบลำดับความสนใจในหมวดหมู่สถานที่ท่องเที่ยวที่แท้จริงกับลำดับที่วิเคราะห์โดยระบบของผู้ใช้ 1 คน

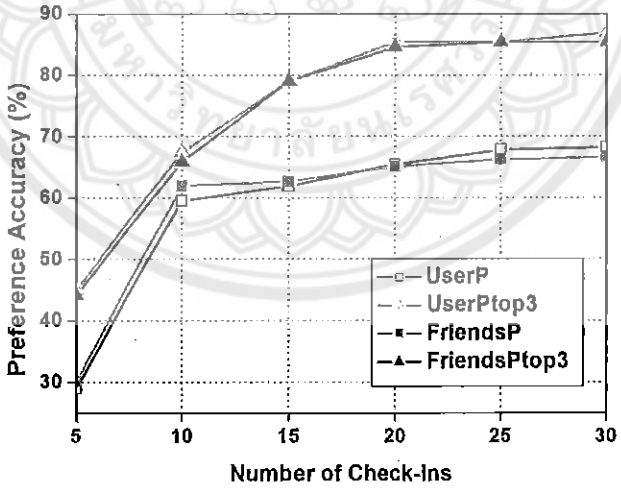
หมวดหมู่สถานที่ท่องเที่ยว	ลำดับความสนใจ		ความถูกต้องในการลำดับความสนใจ
	ข้อมูลที่วิเคราะห์โดยระบบ	ข้อมูลจากผู้ใช้จริง	
สถานที่ท่องเที่ยวทางวัฒนธรรม	5	4	✘
สถานที่ท่องเที่ยวทางประวัติศาสตร์	2	2	✓
สถานที่ท่องเที่ยวเพื่อความบันเทิง	3	3	✓
สถานที่ท่องเที่ยวทางธรรมชาติ	1	1	✓
สถานที่ท่องเที่ยวเพื่อการศึกษา	4	6	✘
สถานที่ท่องเที่ยวทางวิถีชีวิตชุมชน	5	5	✓

ตาราง 14 ลำดับความสนใจในหมวดหมู่สถานที่ท่องเที่ยวที่แท้จริงเปรียบเทียบกับลำดับที่วิเคราะห์จากกลุ่มเพื่อนโดยระบบของผู้ใช้ 1 คน

ลำดับหมวดหมู่สถานที่ท่องเที่ยว	ลำดับความสนใจ		ความถูกต้องในการลำดับความสนใจ
	ข้อมูลที่วิเคราะห์โดยระบบ	ข้อมูลจากกลุ่มเพื่อนผู้ใช้	
สถานที่ท่องเที่ยวทางวัฒนธรรม	5	3	✗
สถานที่ท่องเที่ยวทางประวัติศาสตร์	2	2	✓
สถานที่ท่องเที่ยวเพื่อความบันเทิง	3	4	✗
สถานที่ท่องเที่ยวทางธรรมชาติ	1	1	✓
สถานที่ท่องเที่ยวเพื่อการศึกษา	4	6	✗
สถานที่ท่องเที่ยวทางวิถีชีวิตชุมชน	5	5	✓

โดยแบ่งการทดลองเป็น 2 แบบเพื่อพิจารณาในมิติที่ต่างกัน ดังนี้

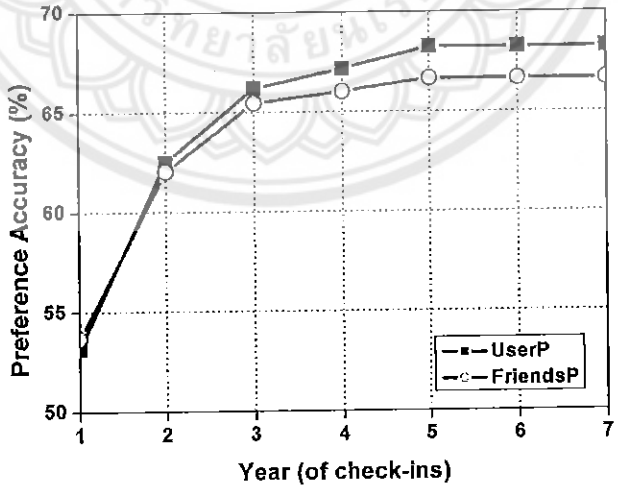
1. พิจารณาความถูกต้องของลำดับความสนใจที่คำนวณจากระบบโดยใช้จำนวนเช็คอินล่าสุดที่ต่างกัน (กำหนดจำนวนเช็คอินล่าสุดเท่ากับ 5 10 15 20 25 30) โดยภาพที่ 17 แสดงค่าความถูกต้องในการลำดับความสนใจกรณีโคลด์สตาร์ทระดับ 1 และ 2



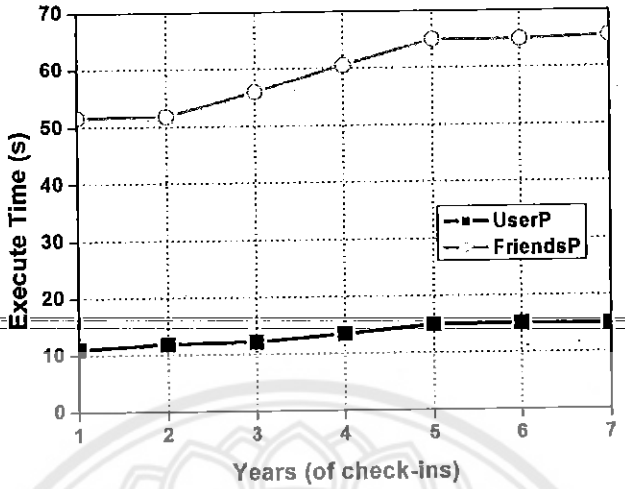
ภาพ 17 ค่าเฉลี่ยความถูกต้องของลำดับหมวดหมู่ความสนใจ

จากการทดลองพบว่าทั้งในกรณีโคลด์สตาร์ทระดับ 1 และ 2 ค่าเฉลี่ยความถูกต้องของการลำดับความสนใจในหมวดหมู่การท่องเที่ยวมีแนวโน้มสูงขึ้นเมื่อมีจำนวนการเช็คอินที่เพิ่มขึ้น โดยโคลด์สตาร์ทระดับ 1 ซึ่งใช้ข้อมูลเช็คอินของผู้ใช้เอง (UserP) มีค่าเฉลี่ยความถูกต้องสูงสุดเท่ากับ 68.25% และมีค่าเฉลี่ยความถูกต้องในการลำดับความสนใจใน 3 อันดับแรกสูงสุด (UserPtop3) เท่ากับ 86.91% เมื่อใช้เช็คอินล่าสุดจำนวน 30 เช็คอิน และโคลด์สตาร์ทระดับ 2 ซึ่งใช้ข้อมูลเช็คอินจากกลุ่มเพื่อนสนิท (FriendsP) มีค่าเฉลี่ยความถูกต้องสูงสุดเท่ากับ 66.67% และมีค่าเฉลี่ยความถูกต้องในการลำดับ 3 อันดับแรกสูงสุด (FriendsPtop3) เท่ากับ 85.40% เมื่อใช้เช็คอินล่าสุดจำนวน 30 เช็คอิน นอกจากนี้ผู้วิจัยได้พิจารณาถึงการใช้จำนวนเช็คอินที่มีผลต่อความถูกต้องเป็นเกณฑ์ในการกำหนดความเพียงพอของข้อมูล โดยให้ค่าความถูกต้องที่ยอมรับได้เป็น 60% ซึ่งจากกราฟในภาพที่ 32 จะให้ค่าเฉลี่ยความถูกต้องน้อยกว่า 60% หากผู้ใช้ที่มีจำนวนการเช็คอินที่ตรวจสอบความถูกต้องแล้วน้อยกว่า 10 ในทั้งสองระดับ ดังนั้นผู้ใช้จะถือว่ามีข้อมูลไม่พอเพียงเมื่อมีจำนวนการเช็คอินน้อยกว่า 10 เช็คอิน

2. พิจารณาความถูกต้องของลำดับความสนใจที่คำนวณจากระบบร่วมกับเวลาที่ใช้ประมวลผลเมื่อใช้ระยะเวลาที่ทำการเช็คอินล่าสุดต่างกัน (กำหนดจำนวนระยะเวลาล่าสุดเท่ากับ 1 2 3 4 และ 5 ปีขึ้นไป) โดยภาพที่ 18 แสดงค่าความถูกต้องในการลำดับความสนใจกรณีโคลด์สตาร์ทระดับ 1 และ 2 และภาพที่ 19 แสดงระยะเวลาที่ใช้ประมวลผลกรณีโคลด์สตาร์ทระดับ 1 และ 2



ภาพ 18 ค่าเฉลี่ยความถูกต้องของลำดับหมวดหมู่ความสนใจเมื่อระยะเวลาเช็คอินต่างกัน



ภาพ 19 ค่าเฉลี่ยเวลาที่ใช้ในการประมวลผลเมื่อระยะเวลาเช็คอินต่างกัน

จากการทดลองพบว่าทั้งในกรณีโคลด์สตาร์ทระดับ 1 และ 2 ค่าเฉลี่ยความถูกต้องของการลำดับความสนใจในหมวดหมู่การท่องเที่ยวสูงขึ้นเมื่อใช้ระยะเวลาที่ทำการเช็คอินที่มากขึ้น โดยโคลด์สตาร์ทระดับ 1 ซึ่งใช้ข้อมูลเช็คอินของผู้ใช้เอง (UserP) มีค่าเฉลี่ยความถูกต้องเท่ากับ 53.15% 62.49% 66.20% 67.16% เมื่อมีระยะเวลาการเช็คอิน 1 2 3 และ 4 ปีตามลำดับ มีค่าเฉลี่ยความถูกต้องเท่ากับ 68.25% เมื่อมีระยะเวลาการเช็คอิน 5 6 และ 7 ปีตามลำดับ และโคลด์สตาร์ทในระดับ 2 ซึ่งใช้ข้อมูลเช็คอินจากกลุ่มเพื่อนสนิท (FriendsP) มีค่าเฉลี่ยความถูกต้องเท่ากับ 53.66% 62.02% 65.46% 66.01% เมื่อมีระยะเวลาการเช็คอิน 1 2 3 และ 4 ปีตามลำดับ มีค่าเฉลี่ยความถูกต้องเท่ากับ 66.67% เมื่อมีระยะเวลาการเช็คอิน 5 6 และ 7 ปีตามลำดับ นอกจากนี้ผู้วิจัยได้พิจารณาถึงเวลาในการประมวลผลพบว่าทั้งในกรณีโคลด์สตาร์ทระดับ 1 และ 2 มีค่าเฉลี่ยเวลาในการประมวลผลมากขึ้นเมื่อใช้ระยะเวลาที่ทำการเช็คอินที่มากขึ้น โดยโคลด์สตาร์ทระดับ 1 มีค่าเฉลี่ยเวลาในการประมวลผลเท่ากับ 10.98 11.86 12.19 13.45 14.98 15.09 14.98 วินาที เมื่อมีระยะเวลาการเช็คอิน 1 2 3 4 5 6และ 7 ปีตามลำดับ และโคลด์สตาร์ทระดับ 2 มีค่าเฉลี่ยเวลาในการประมวลผลเท่ากับ 51.62 51.82 56.02 60.59 64.98 64.98 65.54 วินาที เมื่อมีระยะเวลาการเช็คอิน 1 2 3 4 5 6 และ 7 ปีตามลำดับ อย่างไรก็ตามระยะเวลาที่ใช้ประมวลผลที่ต้องใช้เพิ่มขึ้นเมื่อเทียบกับความถูกต้องในการลำดับความสนใจที่มีความถูกต้องเพิ่มขึ้นแล้ว พบว่ามีความคุ้มค่า เช่น กรณีโคลด์สตาร์ทระดับ 1 สามารถเพิ่มความถูกต้องได้

15.1% โดยใช้ระยะเวลาในการประมวลผลเพิ่มเพียง 4 วินาที และกรณีโคลด์สตาร์ทระดับ 2 สามารถเพิ่มความถูกต้องได้ 13.1% โดยใช้ระยะเวลาในการประมวลผลเพิ่มเพียง13.36 วินาที ผู้วิจัยจึงเลือกข้อมูลเช็คอินจากเช็คอินในระยะเวลา 5 ปี หรือข้อมูลเช็คอินทั้งหมดตลอดระยะเวลาที่ผู้ใช้ใช้งานในเฟสบุ๊คนั่นเอง

นอกจากจำนวนเช็คอินและระยะเวลาในการเช็คอินจะมีผลต่อความถูกต้องในการลำดับความสนใจที่ได้จากผลการทดลองข้างต้นแล้ว ปัจจัยอื่นที่อาจมีผลต่อความถูกต้องในการลำดับความสนใจคือการใช้ข้อมูลจากเครือข่ายสังคมตลอดช่วงเวลา ไม่ได้เลือกช่วงเวลาเฉพาะที่ผู้ใช้ต้องการไปยังสถานที่ท่องเที่ยว ซึ่งอาจทำให้ค่าความถูกต้องสูงขึ้นได้ นอกจากนี้เกณฑ์การเลือกกลุ่มเพื่อนสนิทจากอัลกอริทึม ay-fb-friend-rank (gajus 2012) ที่ไม่ได้นำค่า Time Decay มาใช้ เนื่องจากข้อจำกัดด้านเวลาในการประมวลผลของระบบ ดังนั้นการคำนวณเพื่อเลือกกลุ่มเพื่อนสนิทจึงใช้การพิจารณาปฏิสัมพันธ์ระหว่างกันตลอดช่วงเวลา ไม่ได้เลือกกลุ่มเพื่อนจากที่มีการปฏิสัมพันธ์มากขณะปัจจุบันหรือช่วงเวลาเฉพาะที่ผู้ใช้ต้องการไปยังสถานที่ท่องเที่ยว ซึ่งอาจทำให้ค่าความถูกต้องสูงขึ้นได้

1.3 การทดสอบหาจำนวนของเพื่อนที่เหมาะสม ในกรณีใช้ข้อมูลจากเพื่อนในการหาความสนใจ (Optimal Number of Friends)

นอกจากนี้ในส่วนของการใช้ข้อมูลของเพื่อนเพื่อแก้ไขปัญหาโคลด์สตาร์ทระดับที่ 2 ในการทดสอบที่ 1.2 ผู้วิจัยได้ทดลองถึงการเลือกใช้จำนวนของเพื่อนที่เหมาะสมเพื่อนำมาวิเคราะห์ความสนใจแทนข้อมูลจากผู้ใช้ โดยพิจารณาจากมีความถูกต้องมากที่สุดในการลำดับหมวดหมู่ความสนใจสถานที่ท่องเที่ยว และเวลาใช้เวลาน้อยที่สุดในการประมวลผลเพื่อหาลำดับหมวดหมู่ความสนใจ

เนื่องจากไม่มีวิธีการอย่างเป็นทางการในการเลือกจำนวนเพื่อนที่เหมาะสมข้างต้น ผู้วิจัยจึงใช้วิธี grid search (Cao and Tay 2003) ซึ่งจะกำหนดค่าจำนวนเพื่อนเป็นแบบเอกซ์โพเนนเชียล โดยใช้ค่าตั้งแต่ 2^0 - 2^5 การทดลองทำให้ทราบว่าลำดับความสนใจในหมวดหมู่ของสถานที่ท่องเที่ยวมีลำดับที่ต่างกันเมื่อกำหนดค่าจำนวนเพื่อนเท่ากับ 2^0 - 2^2 แต่จะได้ลำดับความสนใจที่เหมือนกันเมื่อกำหนดค่าจำนวนเพื่อนเท่ากับ 2^3 - 2^5 จากนั้นผู้วิจัยได้ทดลองเพิ่มเติมโดยใช้ค่าจำนวนเพื่อนระหว่าง 2^2 และ 2^3 คือ 5 6 และ 7 ซึ่งผลการทดลองแสดงดัง ตารางที่ 15 โดยการทดสอบของหนึ่งผู้ใช้ จะเห็นว่าลำดับหมวดหมู่ความสนใจสถานที่ท่องเที่ยวมีลำดับที่เหมือนกันเมื่อเลือกใช้ข้อมูลจากเพื่อนจำนวน 5 คน ขึ้นไป เมื่อพิจารณาเวลาที่ใช้ในการประมวลผล จะเห็นว่าเมื่อใช้เพื่อนจำนวนมากขึ้นก็จะใช้เวลาในการประมวลผลนานขึ้น

ตาราง 15 ลำดับความสนใจในหมวดหมู่ของสถานที่ท่องเที่ยวและเวลาในการประมวลผลของหนึ่งผู้ใช้

ลำดับหมวดหมู่สถานที่ท่องเที่ยว	จำนวนเพื่อนที่ใช้คำนวณลำดับหมวดหมู่ความสนใจ								
	2^0	2^1	2^2	5	6	7	2^3	2^4	2^5
สถานที่ท่องเที่ยวทางวัฒนธรรม	5	6	6	6	6	6	6	6	6
สถานที่ท่องเที่ยวทางประวัติศาสตร์	2	2	3	3	3	3	3	3	3
สถานที่ท่องเที่ยวเพื่อความบันเทิง	1	1	1	1	1	1	1	1	1
สถานที่ท่องเที่ยวทางธรรมชาติ	3	3	2	2	2	2	2	2	2
สถานที่ท่องเที่ยวเพื่อการศึกษา	4	4	4	5	5	5	5	5	5
สถานที่ท่องเที่ยวทางวิถีชีวิตชุมชน	5	5	5	4	4	4	4	4	4
ค่าเฉลี่ยเวลาที่ใช้ประมวลผล (วินาที)	13	22	42	61	120	165	176	262	659

1.4 การประเมินประสิทธิภาพในการแนะนำสถานที่ท่องเที่ยว (Recommendation Evaluation)

การทดสอบนี้จะวัดประสิทธิภาพของระบบว่าสามารถแนะนำสถานที่ท่องเที่ยวได้ตรงกับที่ผู้ใช้สนใจหรือค้นหาอยู่ โดยใช้ข้อมูลสถานที่ท่องเที่ยวที่ระบบแนะนำไปยังผู้ใช้จำนวน 10 รายการ (Top-10 Recommendation) เปรียบเทียบกับสถานที่ที่ผู้ใช้มีความสนใจจริง ซึ่งจะใช้วิธีการที่นิยมในระบบแนะนำข้อมูลอย่าง Precision@10 คำนวณได้จากอัตราส่วนของสถานที่ที่ระบบแนะนำและผู้ใช้มีความสนใจต่อสถานที่ที่ระบบแนะนำทั้งหมด ดังสมการที่ 8 นอกจากนี้ยังใช้ Rankscore_{10} ในสมการที่ 9 ซึ่งวัดประสิทธิภาพในการลำดับสถานที่แนะนำ โดยจะมีค่าสูงขึ้นหากผู้ใช้สนใจสถานที่แนะนำในลำดับต้น ๆ

$$precision = \frac{\text{interest} \cap \text{recommended}}{\text{recommended}} \tag{8}$$

$$rankscore = \frac{\sum_{i \in \text{interest} \cap \text{recommended}} \frac{1}{2^{\frac{rank(i)-1}{\alpha}}}}{\sum_{i \in \text{interest}} \frac{1}{2^{\frac{idx(i)-1}{\alpha}}}} \tag{9}$$

การทดสอบใช้ตัวอย่างจำนวน 42 ผู้ใช้ ทั้งในกรณีโคล์ดสตาร์ทระดับที่ 1 ซึ่งแนะนำจากความสนใจของผู้ใช้เอง (User-based) และระดับที่ 2 (Friends-based) ซึ่งแนะนำจากความสนใจของกลุ่มเพื่อนสนิท โดยเปรียบเทียบกับการแนะนำในหลายลักษณะ ซึ่งใช้เกณฑ์ในการ

คำนวณค่าน้ำหนักของสถานที่ในสมการที่ 5 แตกต่างกันไป 5 แบบ คือ

1. Popularity Weight (P) ใช้ค่าความนิยมในสถานที่เพียงอย่างเดียว โดยกำหนดค่า $\beta=0$ และ $\gamma=0$

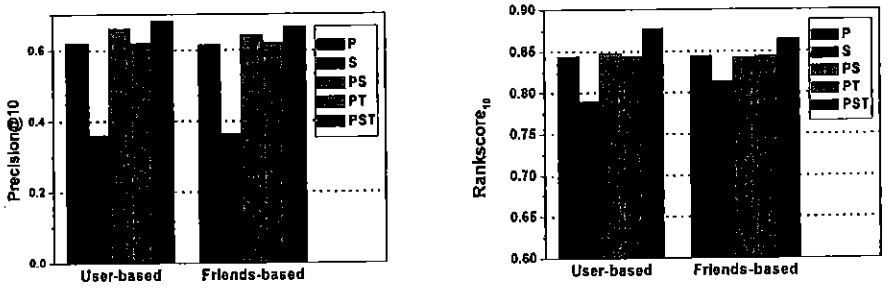
2. Social Weight (S) พิจารณาจากสถานที่ที่เพื่อนสนิทในเฟสบุ๊คของผู้ใช้เคยไป โดยกำหนดค่า $\alpha=0$ และ $\gamma=0$

3. Popularity/Social Weight (PS) พิจารณา Popularity Weight (P) ร่วมกับ Social Weight (S)

4. Popularity/Time Weight (PT) พิจารณา Popularity Weight (P) ร่วมกับค่าช่วงเวลาที่เหมาะสม (Time -T)

5. Popularity/Social/Time Weight (PST) พิจารณาทั้ง 3 ค่าร่วมกัน คือ ค่าความนิยม สถานที่ที่เพื่อนสนิทในเฟสบุ๊คของผู้ใช้เคยไป และค่าช่วงเวลาที่เหมาะสม (Time -T) ในการไปเยี่ยมชมสถานที่

ผลการทดสอบพบว่าการแนะนำทั้งแบบ User-based และ Friends-based การใช้เกณฑ์ในการคำนวณค่าน้ำหนักของสถานที่แบบ PST จะมีค่า Precision โดยเฉลี่ยสูงที่สุด 0.6810 และ 0.6643 ตามลำดับ เช่นเดียวกับค่า Rankscore การใช้เกณฑ์แบบ PST จะให้ค่าเฉลี่ยสูงสุด โดย User-based มีค่าเฉลี่ย Rankscore เท่ากับ 0.8775 และแบบ Friends-based เท่ากับ 0.8647 ดังกราฟในภาพที่ 20 ดังนั้นสถานที่เพื่อนเคยไปและค่าของเวลาที่เหมาะสมในการไปเยี่ยมชมสถานที่ จึงมีผลต่อการเลือกสถานที่ที่ผู้ใช้สนใจหรือค้นหาอยู่ ซึ่งสามารถเพิ่มประสิทธิภาพการแนะนำสถานที่ท่องเที่ยวให้ดีขึ้นได้ อย่างไรก็ตามพบว่าในบางผู้ใช้ การใช้ค่าของเวลาที่เหมาะสมเพิ่มเข้ามาในทั้งแบบ PT และ PTS ทำให้ค่า Precision และ Rankscore ต่ำลง เนื่องมาจากผู้ใช้ระบุช่วงเวลาไม่ตรงกับที่ผู้ใช้สนใจจริง เช่น ระบุช่วงเวลาหน้าร้อนให้กับระบบ แต่ช่วงเวลาจริงที่จะไปเป็นหน้าหนาว ทำให้ระบบแนะนำน้ำตกแทนที่จะเป็นภูเขา นอกจากนี้ในบางตัวอย่างก็ให้ค่า Precision และ Rankscore เท่ากันในการพิจารณาเกณฑ์ทั้งห้าแบบ เนื่องมาจากเพื่อนสนิทในเฟสบุ๊คมีข้อมูลการเยี่ยมชมสถานที่น้อย ไม่ส่งผลต่อการเปลี่ยนแปลงคำแนะนำ รวมถึงผู้ใช้มีความสนใจในสถานที่ไม่มีช่วงเวลาพิเศษ เช่น โบราณสถานหรือวัดบางแห่ง ค่าช่วงเวลาจึงไม่ส่งผลต่อการเปลี่ยนแปลงคำแนะนำเช่นกัน



ภาพ 20 Precision@10 และ Rankscore10 ของการแนะนำเปรียบเทียบระหว่างการใช้เกณฑ์ที่ต่างกัน

บทที่ 5
บทสรุป

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้พัฒนาเทคนิคการแก้ปัญหาของระบบแนะนำข้อมูลสำหรับนักท่องเที่ยว โดยใช้ประโยชน์จากเครือข่ายสังคมออนไลน์ ได้แก่ เฟสบุ๊ค ซึ่งเป็นข้อมูลที่มีประโยชน์ในหลายๆ ด้าน เช่น สามารถนำไปวิเคราะห์ในการแนะนำสินค้า หรือบริการต่างๆ ในเชิงธุรกิจได้ ผลการทดสอบแสดงให้เห็นถึงความน่าเชื่อถือ และประโยชน์ที่ได้จากข้อมูลดังกล่าว โดยแนวคิดใหม่ของงานวิจัยนี้คือกระบวนการในการแก้ปัญหาโคล์สตาร์ท ที่แตกต่างไปจากงานวิจัยอื่นๆ ถึงแม้จะมีงานวิจัยหลายๆ ชิ้นที่ใช้ข้อมูลในการเช็คอินสถานที่ต่างๆ เหมือนกันก็ตาม

ผลการทดสอบความพึงพอใจของผู้ใช้ระบบสรุปได้ว่า ผลการดำเนินงานเป็นไปตามแผนการดำเนินงานวิจัย มีความสอดคล้องกับวัตถุประสงค์และสมมติฐานของการวิจัย โดยมีผลประเมินความพึงพอใจในการใช้ระบบด้านส่วนติดต่อผู้ใช้งานและการแนะนำอยู่ในระหว่างระดับดีและดีมาก ผู้ใช้มีแนวโน้มที่จะใช้แอปพลิเคชันในการช่วยหาข้อมูลการท่องเที่ยวต่อไปและแนะนำแอปพลิเคชันให้กับบุคคลอื่นต่ออีกด้วย นอกจากนี้ผู้วิจัยยังได้นำข้อเสนอแนะและข้อคิดเห็นในการประเมินโครงการวิจัยในแต่ละระยะจากทั้งผู้ทรงคุณวุฒิภายในและนอกมหาวิทยาลัย มาทำการปรับปรุงและพัฒนาระบบให้มีประสิทธิภาพและตอบสนองการใช้งานของผู้ใช้ให้สมบูรณ์มากยิ่งขึ้น

5.2 ข้อเสนอแนะ

การพัฒนาในขั้นต่อไปของงานวิจัยนี้คือการวิเคราะห์ความสนใจจากเครือข่ายสังคมโดยแบ่งพิจารณาเป็นช่วงเวลา เช่น ความสนใจสถานที่ท่องเที่ยวในฤดูที่แตกต่างกันเป็นต้น เพื่อเพิ่มประสิทธิภาพในการคาดการณ์ความสนใจของผู้ใช้ให้มีความถูกต้องมากยิ่งขึ้น ซึ่งอาจจะทำให้การแนะนำข้อมูลการท่องเที่ยวตรงกับความสนใจและความต้องการของผู้ใช้มากขึ้นด้วย นอกจากนี้ยังอาจจะมีการตรวจสอบความสนใจที่เปลี่ยนไปของผู้ใช้แต่ละคน ทั้งนี้เนื่องจากความสนใจของผู้ใช้อาจจะมีการเปลี่ยนแปลงได้ตลอดเวลา ดังนั้นการแนะนำแบบเดิมๆ อาจจะทำให้การแนะนำไม่ตรงกับความสนใจของผู้ใช้ได้

บรรณานุกรม

ไกรศักดิ์ เกษร. (2555). ระบบค้นคืนสารสนเทศ: แนวคิดและแนวทางการพัฒนาในอนาคต. พิษณุโลก: โฟกัสพริ้นติ้ง.

นฤพนธ์ พนาวงศ์ และจักรกฤษณ์ เสนห์. (2553). ระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทยด้วยหลักการออนโทโลจีและเนมแมทชิ่ง. Journal of Information Science and Technology, 1(2), 60–69.

Abel, F., Gao, Q., Houben, G.-J. and Tao, K. (2011). Analyzing User Modeling on Twitter for Personalized News Recommendations. In J. A. Konstan, R. Conejo, J. L. Marzo and N. Oliver (Eds.), User Modeling, Adaption and Personalization (p. 1–12). Heidelberg: Springer-Verlag.

Arhippainen, L., Rantakokko, T. and Tähti, M. (2004). Mobile Feedback Application for Emotion and User Experience Collection. In Proceedings of Proactive Computing Workshop (pp. 77–81). Helsinki: Helsinki University Press.

Burke, R. (2000). Knowledge-Based Recommender Systems. In Encyclopedia of Library and Information (p. 2000). New York: Marcel Dekker.

Castells, P., Fernandez, M. and Vallet, D. (2007). An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. IEEE Transactions on Knowledge and Data Engineering, 19(2), 261–272.

Cheverst, K., Mitchell, K. and Davies, N. (2002). The Role of Adaptive Hypermedia in a Context-aware Tourist GUIDE. Communications of the ACM - The Adaptive Web, 45(5), 47–51.

Chin, D. N. and Porage, A. (2001). Acquiring User Preferences for Product Customization. In M. Bauer, P. J. Gmytrasiewicz and J. Vassileva (Eds.), User Modeling 2001 (pp. 95–104). London: Springer-Verlag.

Coyle, L. and Cunningham, P. (2003). Exploiting Re-ranking Information in a Case-Based Personal Travel Assistent. In Workshop on Mixed-Initiative Case-Based Reasoning at the 5th International Conference on Case-Based Reasoning (pp. 11-20).

Firan, C. S., Nejdl, W. and Paiu, R. (2007). The Benefit of Using Tag-Based Profiles. In Proceedings of the 2007 Latin American Web Conference (pp. 32–41). Washington, DC, USA: IEEE Computer Society.

gajus. (November 16, 2012). ay-fb-friend-rank. Retrieved January 12, 2014, from https://github.com/gajus/facebook-friend-rank

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R. and Gómez-Berbís, J. M. (2009). SPETA: Social pervasive e-Tourism advisor. Telematics and Informatics, 26(3), 306–315.

Hang, Y., Guiran, C. and Xingwei, W. (2009). A Cold-Start Recommendation Algorithm Based on New User's Implicit Information and Multi-attribute Rating Matrix. In Ninth International Conference on Hybrid Intelligent Systems (pp. 353–358). Shenyang: IEEE.

Harfield, A. (2014). Tourism Database for lower north of Thailand. Retrieved January 17, 2014, from http://www.mobcomlab.com

Kabassi, K. (2010). Personalizing recommendations for tourists. Telematics and Informatics, 27(1), 51–66.

Mezhoudi, N. (2013). User Interface Adaptation Based on User Feedback and Machine Learning. In Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces Companion (pp. 25–28). New York: ACM.

Moreno, A., Valls, A., Isern, D., Marin, L. and Borràs, J. (2013). SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities. Engineering Applications of Artificial Intelligence, 26(1), 633–651.

Rashid, A. M., Karypis, G. and Riedl, J. (2008). Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach. SIGKDD Explorations Newsletter, 10(2), 90–100.

Widman, J. (2014). EdgeRank. Retrieved July 2, 2014, from http://edgerank.net/

Widyantoro, D. H., Ioerger, T. R. and Yen, J. (1999). An Adaptive Algorithm for Learning Changes in User Interests. In Proceedings of the Eighth International Conference on Information and Knowledge Management (pp. 405–412). New York: ACM.

Zhu, Y., He, L. and Wang, X. (2012). User Interest Modeling and Self-Adaptive Update Using Relevance Feedback Technology. Procedia Engineering, 29, 721–725.