สถาปัตยกรรมระบบเพื่อการจัดการข้อมูลบนบริการเครือข่ายทางสังคม

ปิติ จำปีทอง[1] และ วีระศักดิ์ คุรุธัช[2]

คณะวิทยาการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีมหานคร กรุงเทพฯ

Emails: pitichampeethong@gmail.com[1], werasak@mut.ac.th[2]

บทคัดย่อ

บทความฉบับนี้นำเสนอ สถาปัตยกรรมระบบ ที่ใช้ในการเก็บรวบรวมข้อมูลบน Social Network Service (SNS) จากหลายๆแหล่ง มารวบรวมและจัดเก็บเป็นคลังข้อมูล เพื่อนำข้อมูลที่ได้ไปใช้ในการวิเคราะห์พฤติกรรมในด้านต่างๆ เช่น การวิเคราะห์ข้อมูลผู้บริโภคเพื่อช่วยพัฒนาสินค้า หรือการดูแนวโน้มของสินค้าในตลาด เป็นต้น บทความฉบับนี้ช่วยแก้ปัญหาใน 2 ประเด็น ประเด็นแรก คือ การแก้ปัญหาการถูกจำกัดเวลาในการเก็บรักษาข้อมูลบน SNS และอีกประเด็น คือ การแก้ปัญหาในเรื่องของข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลา ยากต่อการค้นหาข้อมูลในสิ่งที่ผู้ใช้สนใจ ผลลัพธ์ที่ได้จากบทความฉบับนี้ คือ ข้อมูลจาก SNS หลายแหล่งมาเก็บไว้ได้ยาวนานขึ้น เพื่อนำข้อมูลไปวิเคราะห์ในสิ่งที่ผู้ใช้สนใจ ซึ่งข้อมูลที่ได้จะมีความน่าเชื่อถือและเป็นประโยชน์มากยิ่งขึ้น

คำสำคัญ-- Social Network Service; Social Data Analysis; Web Intelligent

1. บทนำ

ในปัจจุบัน Social Network Service (SNS) เป็นแหล่งข้อมูลที่สำคัญมากยิ่งขึ้น เนื่องจากมีผู้ใช้บริการเป็นจำนวนมาก ซึ่งข้อมูลที่ผู้ใช้ได้ทำการบันทึกไว้บน SNS นั้น อาจจะเป็นข้อมูลที่จะสามารถช่วยองค์กร ในเรื่องของการวิเคราะห์ข้อมูลในด้านต่างๆ เช่น การส่งเสริมการขาย, การพัฒนาผลิตภัณฑ์ หรือ การติดตามความเคลื่อนไหวของสินค้าของคู่แข่งขัน เป็นต้น ยกตัวอย่าง SNS ที่มีผู้ใช้งานอยู่เป็นจำนวนมากเช่น Twitter.com, Facebook.com และ LinkedIn.com เป็นต้น ซึ่งส่วนใหญ่เว็บดังกล่าวมีระบบการสมัครสมาชิก และมีการจัดการในเรื่องของสิทธิความเป็นส่วนตัว (Privacy) เพื่อป้องกันการเข้าถึงข้อมูลที่ไม่อนุญาตให้เปิดเผยได้ ก่อนที่จะได้รับอนุญาตจากสมาชิกคนนั้นก่อน

มีบทความที่กล่าวถึงการใช้งานบน SNS อยู่เป็นจำนวนไม่น้อย โดยบทความส่วนใหญ่นั้นจะกล่าวถึงการนำเครื่องมือ การวิเคราะห์เครือข่ายทางสังคม หรือ Social Network Analysis (SNA) มาใช้ทำการวิเคราะห์ข้อมูลพื้นฐานของสมาชิก ตัวอย่างข้อมูลพื้นฐาน เช่น ชื่อ อายุ หรือวันเดือนปีเกิด ในการค้นหาความสัมพันธ์ของแต่ละสมาชิก เพื่อแบ่งกลุ่มของสมาชิกที่มีความสัมพันธ์คล้ายกันตามข้อมูลพื้นฐาน และมีบางบทความได้นำเสนอเทคนิคการทำคาต้าไมนิ่งบนเว็บ (Web Mining Technique) มารวมกับข้อมูลที่ได้จากการใช้เครื่องมือ SNA เพื่อสร้างระบบการค้นหาข้อมูล ซึ่งบทความที่ได้กล่าวมาข้างต้นนั้น ส่วนใหญ่ใช้ข้อมูลในการทดสอบที่เป็นข้อมูลแบบมีโครงสร้างที่แน่นอน (Structured Data) เนื่องจากได้สร้างระบบเพื่อใช้ในการทดสอบขึ้นมาเอง ดังนั้นจึงสามารถกำหนดรูปแบบโครงสร้างของข้อมูลได้ หรือบางบทความได้ใช้ระบบที่ไม่มีการจัดการในเรื่องของ สิทธิความเป็นส่วนตัว (Privacy) มาทดสอบ ซึ่งข้อมูลทดสอบที่กล่าวมาจะแตกต่างกับข้อมูลที่ใช้ในการทดสอบของบทความฉบับนี้ เนื่องจากข้อมูลที่ใช้ในการการทดสอบ มาจาก SNS ที่เป็นระบบปิด และมีโครงสร้างของข้อมูลที่ไม่แน่นอนเพราะว่าขึ้นอยู่กับโครงสร้างของ SNS แต่ละแหล่ง

จากที่กล่าวมาข้างต้น บทความฉบับนี้ได้นำเสนอสถาปัตยกรรมระบบ ที่สามารถเก็บรวบรวมข้อมูลจาก SNS หลายแหล่งมาเก็บไว้ เพื่อนำข้อมูลที่ได้จากการเก็บรวบรวมไปวิเคราะห์ และแสดงผลลัพธ์ให้ง่ายต่อการใช้งานของผู้ใช้ต่อไป

โดยข้อมูลที่ใช้ในการทดสอบสถาปัตยกรรมระบบ ดังที่กล่าวมาข้างต้นนั้น บทความฉบับนี้เลือกใช้ข้อมูลจาก Facebook.com (FB) เป็นกรณีศึกษา โดยวิธีการเก็บรวบรวมข้อมูลจาก FB นั้นได้เลือกใช้เทคโนโลยี PHP ร่วมกับเทคโนโลยี Asynchronous JavaScript And XML (AJAX) ในการติดต่อไปยัง FB ผ่านทางชุดคำสั่งเฉพาะที่เรียกว่า Facebook Connect เพื่อนำข้อมูลของแต่ละสมาชิกมาจัดเก็บลงใน ฐานข้อมูล โดยใช้เทคนิคฟังก์ชั่นเรียกตัวเอง (Recursive function) หลังจากนั้น ได้แสดงผลลัพธ์ของข้อมูลที่ได้จัดเก็บในฐานข้อมูล ด้วยการใช้อัลกอริธึมในการแบ่งกลุ่มของข้อมูลการสนทนาที่มีหัวข้อเดียวกัน จัดอยู่ในกลุ่มเดียวกัน ซึ่งผลลัพธ์ที่ได้จะอยู่ในรูปแบบที่ง่ายต่อผู้ใช้ เพื่อที่จะนำข้อมูลเหล่านั้นไปวิเคราะห์เป็นเป็นประโยชน์ต่อไป

2. ที่มาและแรงจูงใจของปัญหา

จากการที่มีผู้ใช้บริการบน SNS ที่เพิ่มขึ้นเป็นจำนวนมาก เช่น Facebook.com, Twitter.com และ Linkedin.com ซึ่งเว็บไซต์ดังกล่าวจะอนุญาตให้ผู้ที่สมัครเป็นสมาชิก สามารถแบ่งปันข้อมูล หรือข่าวสารต่างๆ ของสมาชิกเอง ไปยังกลุ่มเพื่อนสมาชิกด้วยกัน ทำให้ SNS กลายเป็นแหล่งข้อมูลที่น่าสนใจอย่างมาก เพื่อที่จะนำข้อมูลเหล่านั้น มาวิเคราะห์ในมุมมองด้านต่าง เช่น เพื่อดูแนวโน้ม หรือ ความเคลื่อนไหวต่างๆ ของสินค้า ผู้บริโภค หรือคู่แข่งขันได้

แต่เนื่องจากระบบ SNS ส่วนใหญ่นั้น ถูกออกแบบมาให้มีการจัดการเรื่องสิทธิความเป็นส่วนตัว (Privacy) ของแต่ละสมาชิก ซึ่งยากต่อการเข้าถึงข้อมูลต่างๆของแต่ละสมาชิก ซึ่งหมายความว่าไม่อนุญาตให้เข้าถึงข้อมูลของแต่ละสมาชิกได้ ถ้าหากไม่ได้รับการอนุญาตจากสมาชิกคนนั้นก่อน และข้อมูลของสมาชิกแต่ละคนที่เผยแพร่ออกมานั้นมีการจำกัดเวลาในการแสดงผลข้อมูล และมีการเปลี่ยนแปลงตลอดเวลา เนื่องจากพื้นที่ในการแสดงข้อมูลนั้นมีอยู่อย่างจำกัด จึงทำให้โอกาสที่ข้อมูลที่เผยแพร่บน SNS นั้น มีโอกาสที่จะสูญหายไปโดยที่ยังไม่ได้รับการเก็บรวบรวม ซึ่งอาจจะมีบางข้อมูลเป็นประโยชน์และ มีความสำคัญต่อการนำไปวิเคราะห์ จากที่กล่าวมานั้น จึงสามารถจะสรุปปัญหาที่น่าสนใจดังนี้

1) ข้อจำกัดเรื่องของเวลาที่จะรักษาข้อมูลไว้บน SNS
2) การเข้าถึงข้อมูลความเป็นส่วนตัวของสมาชิก
3) ข้อมูลจำนวนมากที่ไม่ได้รับการนำมาใช้ประโยชน์
4) ข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็วทำให้ยากต่อการค้นหา

จากปัญหาที่กล่าวมาข้างต้นนั้น บทความฉบับนี้ ได้นำเสนอสถาปัตยกรรมระบบ ที่มีวิธีการเก็บรวบรวมข้อมูล โดยการติดต่อไปยัง SNS ผ่านทางชุดคำสั่งเฉพาะ (Application Programming Interface - API) ที่แต่ละแหล่งได้เปิดให้ใช้ เพื่ออนุญาตให้เก็บรวบรวมข้อมูลของแต่ละสมาชิกที่เผยแพร่ได้ แล้วนำข้อมูลที่ได้เก็บไว้บนฐานข้อมูล และนำเสนอข้อมูลที่เก็บไว้ออกมาอยู่ในรูปแบบที่ง่ายต่อผู้ใช้ เพื่อนำไปวิเคราะห์ให้เป็นประโยชน์ต่อไป

3. งานและทฤษฎีที่เกี่ยวข้อง

3.1. Social Network Analysis (SNA)

เป็นการวิเคราะห์ข้อมูลพื้นฐาน ตัวอย่างเช่น ชื่อ อายุ หรือวันเดือนปีเกิดของแต่ละสมาชิก เพื่อค้นหาความสัมพันธ์ของสมาชิกทั้งหมดบนเครือข่ายทางสังคัมเดียวกัน โดยใช้เครื่องมือ SNA ในการวิเคราะห์ และนำข้อมูลที่ได้จากการใช้เครื่องมือ SNA มาใช้ร่วมกับเทคนิคของการดาต้าไมนิ่งบนเว็บ เพื่อสร้างระบบการค้นหาข้อมูล ที่สามารถแสดงข้อมูลบนเว็บของแต่ละสมาชิก ที่มีความสัมพันธ์ตรงกันกับสิ่งที่ผู้ใช้สนใจ [1], [2] โดยสิ่งที่แตกต่างจากบทความฉบับนี้ คือข้อมูลที่ได้ทำการทดสอบนั้น เป็นข้อมูลที่มีการกำหนดรูปแบบโครงสร้างที่แน่นอน เนื่องจากสร้างระบบขึ้นมาใช้ในการทดสอบเอง อีกทั้งข้อมูลที่ใช้ในการทดสอบนั้น ยังอยู่ในรูปแบบของข้อความ (Text) เพียงอย่างเดียว แต่กับบทความฉบับนี้ สามารถดึงข้อมูลที่อยู่ในรูปแบบของสื่อ ที่หลากหลายชนิด เช่น ข้อความ เอกสารรูปภาพ หรือไฟล์เสียง จากระบบ SNS ที่มีระบบรักษาสิทธิความเป็นส่วนตัวของข้อมูลของสมาชิกได้

3.2. Social network based on mobile communication

[3] ได้อธิบายวิธีการวิเคราะห์ข้อมูล ที่ได้จากการเก็บข้อมูลการสนทนาระหว่างผู้ใช้แต่ละราย ผ่านทางอุปกรณ์มือถือ ซึ่งจะเห็นได้ว่ามีความคล้ายกับบทความฉบับนี้ ผลลัพธ์ที่ได้ของ [3] คือ เครือข่ายทางสังคม (Social network) ที่สามารถบอกถึงความสัมพันธ์ของผู้ใช้มือถือแต่ละรายได้

3.3. Social network based on email communication

[4] ได้อธิบายวิธีการสร้างเครือข่ายทางสังคม (Social network) โดยการวิเคราะห์ข้อมูลจากเนื้อหาที่มีอยู่ในอีเมล์ของผู้ใช้แต่ละราย เพื่อใช้ในการจัดระดับความสำคัญของอีเมล์แต่ละฉบับของผู้ใช้แต่ละราย ว่ามีลำดับมากปานกลาง หรือน้อยม เมื่อมีการส่งอีเมล์ไปยังผู้รับปลายทาง

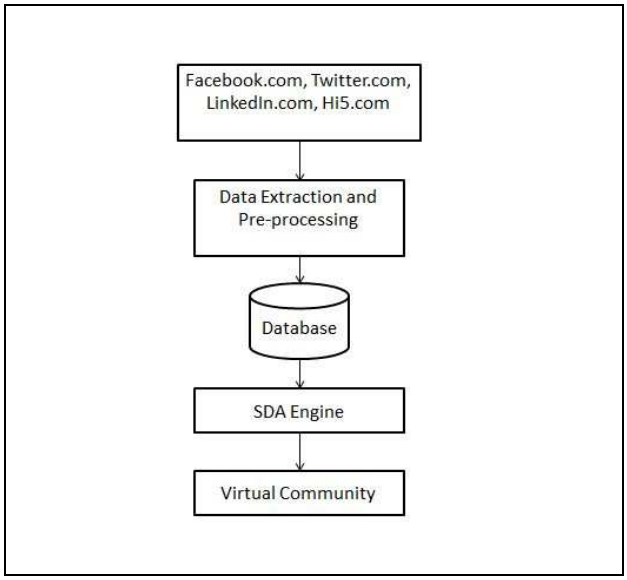
3.4. Social network based on instant messenger

[5] ได้อธิบายการสร้าง การจัดลำดับความสำคัญของสมาชิกที่สนทนา ผ่านทางเว็บไซต์ Tlen.pl โดยการวิเคราะห์ข้อมูลที่ได้จากการสนทนา เพื่อค้นหาความสัมพันธ์ของแต่ละสมาชิก โดยเลือกใช้การเชื่อมโยงไปยัง Tlen.pl ผ่านทางชุดคำสั่งพิเศษ เพื่อเก็บข้อมูลการสนทนาของแต่ละสมาชิก ซึ่งจะคล้ายกับบทความฉบับนี้ แต่ต่างกันตรงที่ [5] เน้นไปที่การค้นหาความสำคัญของสมาชิกที่สนทนาร่วมกัน แต่กับบทความนี้เน้นการจัดการข้อมูลเพื่อให้ง่ายต่อการนำไปใช้งาน และเก็บรวบรวมจากหลายๆแหล่ง

3.5. Social Tagged Visualization

[6], [7] เป็นการนำเสนอวิธีการจัดกลุ่มของป้ายชื่อ หรือ Tag โดยการค้นหาความสัมพันธ์จากข้อมูลของผู้ใช้แต่ละราย บนเว็บไซต์ต่างๆ กับ Tag ที่ผู้ใช้ได้ทำการเลือกเอาไว้เพื่อนำมาสร้างกลุ่มของ Tag ที่มีความเกี่ยวข้องกัน เพื่อสะดวกต่อผู้ใช้ในการค้นหาข้อมูลตามกลุ่มของ Tag ในเรื่องที่ผู้ใช้สนใจได้ถูกต้องและแม่นยำยิ่งขึ้น

4. ระบบที่นำเสนอ

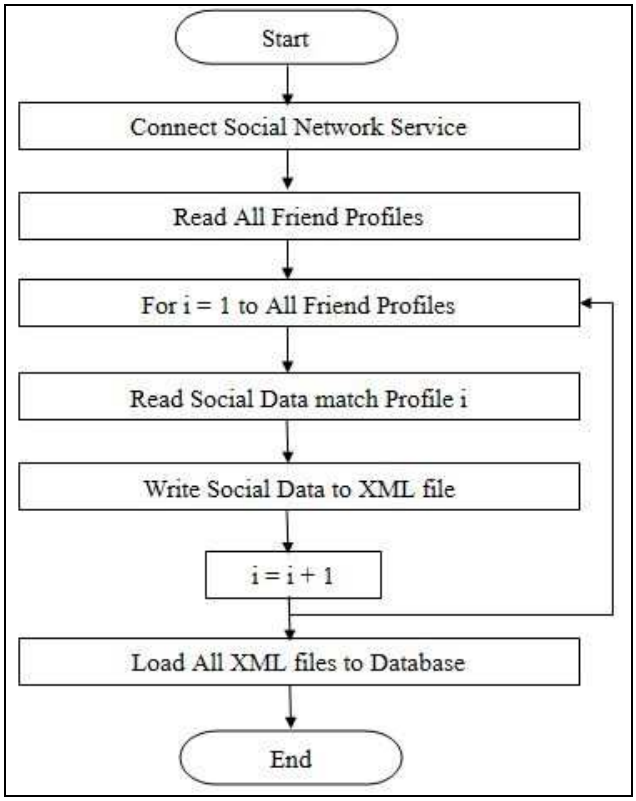


รูปที่ 1. ภาพรวมของระบบ

4.1. Data Extraction and Pre-processing

เป็นขั้นตอนที่ทำการติดต่อไปยัง SNS ในแต่ละแหล่ง โดยผ่านทางชุดคำสั่งเฉพาะที่ SNS แต่ละแหล่งได้จัดเตรียมเอาไว้ ในที่นี้ได้ทดสอบไปติดต่อยัง Facebook.com ผ่านทางชุดคำสั่งเฉพาะคือ Facebook Connect ซึ่งเมื่อทำการติดต่อได้แล้ว ก็จะสามารถที่จะดึงข้อมูลของเพื่อนสมาชิกได้ซึ่งข้อมูลที่ได้ในขั้นตอนนี้จะบันทึกเก็บไว้อยู่ในรูปของเอกสาร XML ดังรูปที่ 2.

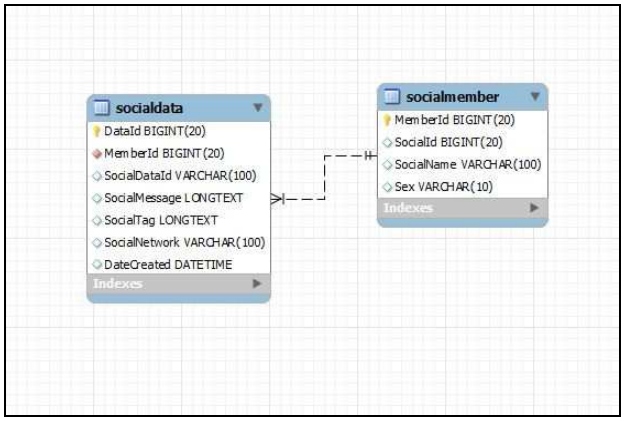
เนื่องจากมีข้อจำกัดเรื่องของเวลาที่อนุญาตให้เข้าถึงข้อมูลบน SNS นั้น จึงต้องแบ่งการติดต่อ ออกเป็นสองส่วน ในการติดต่อไปดึงข้อมูลของเพื่อนสมาชิกแต่ละรายคือ ส่วนที่หนึ่ง ติดต่อไปดึงข้อมูลพื้นฐานของสมาชิกแต่ละรายก่อน ตัวอย่างข้อมูลพื้นฐานของสมาชิก เช่น ชื่อ นามสกุล เพศ และวันเดือนปีเกิด หลังจากนั้นจึงติดต่อไปยังอีกส่วนคือ ข้อมูลที่สมาชิกแต่ละรายได้แบ่งปันไว้บนหน้าแรกของระบบสมาชิกรายนั้น โดยเลือกใช้เทคนิคในการติดต่อที่เรียกว่า ฟังก์ชั่นเรียกตัวเอง เมื่อได้ข้อมูลทั้งหมดมาแล้วก็จะทำการคัดแยกข้อมูลต่างๆเพื่อจัดเก็บลงในฐานข้อมูลต่อไป



รูปที่ 2. Data Extraction and Pre-processing

4.2. Database

ในขั้นตอนนี้หลังจากที่ได้ข้อมูลมาอยู่ในรูปแบบเอกสาร XML แล้ว ก็ได้ทำการคัดแยกข้อมูล เพื่อแยกข้อมูล XML เหล่านั้นบันทึกลงไปในฐานข้อมูลที่ได้ทำการออกแบบเอาไว้ ใน รูปที่ 3. เพื่อจะนำข้อมูลทั้งหมดเก็บเอาไว้ใช้ เพื่อการวิเคราะห์ต่อไป



รูปที่ 3. ฐานข้อมูลของระบบ

4.3. Social Data Analysis Engine (SDA Engine)

เป็นขั้นตอนการนำข้อมูลที่มีอยู่ในฐานข้อมูล มาค้นหาความสัมพันธ์ของของสมาชิก โดยนำข้อมูลของสมาชิกกับข้อมูลที่เผยแพร่ของสมาชิกแต่ละราย มาวิเคราะห์เพื่อให้ได้ข้อมูลที่ตรงกับความต้องการของผู้ใช้ โดยใช้เทคนิคในการร้องขอข้อมูล (Query) ในแบบต่างๆ รวมถึงใช้เทคนิคการทำดาต้าไมนิ่ง เพื่อให้ได้ข้อมูลตามที่ผู้ใช้ต้องการ

4.4. Virtual Community

เป็นการแสดงผลลัพธ์ที่ได้จากขั้นตอนของ SDA Engine มาแสดงให้อยู่ในรูปแบบที่สามารถจะแนะนำผู้ใช้ หรือ/และเป็นเครื่องมือช่วยเหลือผู้ใช้ ในการค้นหา หรือวิเคราะห์ข้อมูล เพื่อสนับสนุนการตัดสินใจ และเป็นประโยชน์ต่อผู้ใช้

5. การทดสอบการใช้งาน

5.1. เครื่องมือที่ใช้ในการทดสอบ

1) Zend Server Community Edition
2) jQuery Framework
3) MySQL 5.0
4) Facebook Connect

5.2. อัลกอริธึมของการแบ่งกลุ่มข้อมูลที่มีความคล้ายกัน

ในขั้นตอนของ SDA Engine ได้นำเสนอการนำป้ายชื่อ หรือ Tag ที่ได้จากการเก็บข้อมูลจาก Facebook.com มาผ่านอัลกอริธึมนี้ เพื่อสร้างระดับความสำคัญให้แก่ ป้ายชื่อ หรือ Tag นั้นๆ ดังรูปที่ 4. และมีรายละเอียดของอัลกอริธึมดังนี้

```
READ ItemList from Database
SET MinSize to 10
SET MaxSize to 30
SET SizeOfFont to 0
SET MaxValue to value of first position in ItemList
SET MinValue to value of last position in ItemList
SET DiffValue to (MaxValue - MinValue)
IF DiffValue = 0 THEN
   SET DiffValue to 1
END IF
SET StepValue to (MaxSize - MinSize) / DiffValue
FOR each Item in ItemList
  SET Keyword to Tag in Item
  SET Frequency to Value in Item
  SET DiffFreq to (Frequency - MinValue)
  SET SizeOfFont to Round ((MinSize + DiffFreq) * StepValue)
  DISPLAY SizeOfFont
END FOR
```

FarmVille	Twitter	Quiz Planet!	Twitter"->ผ่าน พาจ Twitter	Zoo World	Barn Buddy	Restaurant City	Facebook for iPhone	Happy Aquarium	Mafia Wars
Love Gifts♥	Tumblr	Café World	RSS Graffiti	Social City	Treasure Isle	Astrology	Pet Society	Quiz Creator	Pillow Fight
Friends For Sale!	Facebook for iPhone"->ใช้ Facebook for iPhone	SPP Ranch!	Your Look (daily)	Application Builder (Create Your Own App)	Facebook for iPhone"->ผ่านทาง Facebook for iPhone	Food Fight	Hotel City	Details	Lover of the day
Hearts♥	Facebook สำหรับ BlackBerry"->ผ่านทาง Facebook สำหรับ BlackBerry	FishVille	Hearts	Mobile Web	Hair Chic donations	Treasure Madness	Country Story	Auto predictions	
Gift Creator	Facebook for BlackBerry® smartphones"->ผ่านทาง Facebook for BlackBerry® smartphones	Dejected Bits	I♡Chocolate	Facebook for BlackBerry® smartphones"->ใช้ Facebook for BlackBerry® smartphones	เว็บสำหรับสื่อสาร"->ผ่านทางเว็บสำหรับสื่อสาร	Cats	School of Wizardry	Love	Sending Love
YouTube	My Gifts	Birthday Cards	Daily Gifts	Happy Island	Medical Mayhem	Facebook Mobile"->ใช้ Facebook Mobile	cinepeopia	Fortune Cookie	Facebook for BlackBerry
♫Happiness♫	Text Message	Sweetheart Gifts♥	iHeart	FarmVille"->ผ่านทาง FarmVille	Texas Hold'Em Poker	Jax Paradise	Super Cupcakes	Smiley♥	Feed vill 3.2
Chocolates	Gifts♡	Give Hearts	SuperPoke! Pets	Castle Age	Friendship Gifts	Warm Gifts	Mafia Wars Weekly Loot Collection	Heart Frenzy❤	Facebook ผ่านเว็บไซต์"->ผ่านทาง Facebook ผ่านเว็บไซต์
SuperPoke	เพื่อน เพื่อน ใช่	Birthday Album	CAKES	Feed vill 3.2"->ผ่านทาง Feed vill 3.2	สำหรับเว็บมาสเตอร์ของเครือข่ายสังคม	Flowers for you!	Smiles	Love & Friendship	Feel Master
คอมไม่เอาไม่เอาเรื่องการใช้งาน??	Kitty Wishes	โจทย์น่าสนใจมีอะไรใหม่!!!

[6] J. Bae, and K. Lee, “TagReel: A Visualization of Tag Relations among User Interests in the Social Tagging System,” Aug. 2009.

[7] A. Zubiaga, A. P. Garcia-Plaza, V. Fregno, and R. Martinez, “Content-based Clustering for Tag Cloud Visualization,” Jul. 2009.