**ข้อเสนอโครงงานมหาบัณฑิต**

**(MASTER PROJECT PROPOSAL)**

**ชื่อเรื่อง (ภาษาไทย)** การปรับปรุงวิธีการจัดลำดับการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่: กรณีศึกษาแอนดรอยด์มาร์เก็ต

ชื่อเรื่อง (ภาษาอังกฤษ) Improving Ranking of Mobile Application Retrieval : Android Market Case Study

เสนอโดย นายอนุกุล ลีลาภัทรกิจ

**เลขประจำตัวนิสิต** 5370334821

**สาขาวิชา** วิศวกรรมซอฟต์แวร์

**ภาควิชา** วิศวกรรมคอมพิวเตอร์

**คณะ** วิศวกรรมศาสตร์

**สถานที่ติดต่อ** 74 ซอยแม้นศรี 2 ถนนบำรุงเมือง แขวงคลองมหานาค   
 เขตป้อมปราบฯ กรุงเทพมหานคร 10100

**โทรศัพท์** 084-333-7626

**อีเมล์** Anugoon.L@Student.chula.ac.th

**อาจารย์ที่ปรึกษา**  ผศ.นครทิพย์ พร้อมพูล

**คำสำคัญ (ภาษาไทย)** การค้นคืนข้อมูล, การค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่,

การจัดลำดับ

**คำสำคัญ (ภาษาอังกฤษ)** Information Retrieval, Mobile Application Retrieval, Ranking

# ที่มาและความสำคัญของปัญหา

ในปัจจุบันจำนวนผู้ใช้อุปกรณ์เคลื่อนที่ เช่น โทรศัพท์เคลื่อนที่แบบสมาร์ทโฟนและอุปกรณ์แท็บเล็ตคอมพิวเตอร์มีอัตราการเติบโตขึ้นอย่างรวดเร็ว ทั้งนี้เนื่องมาจากการที่อุปกรณ์เคลื่อนที่เหล่านี้สามารถตอบสนองการใช้งานได้หลากหลาย และสามารถพกพาไปได้ทุกที่ ด้วยอัตราการเติบโตนี้เองจึงทำให้ผู้พัฒนาซอฟต์แวร์เป็นจำนวนมากมีความสนใจที่จะพัฒนาซอฟต์แวร์ประยุกต์ให้กับอุปกรณ์เคลื่อนที่เหล่านี้ ซึ่งจากผลสำรวจเมื่อวันที่ 7 กรกฎาคม พ.ศ. 2554 จำนวนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ ที่ทำงานบนระบบปฏิบัติการ Apple iOS พบว่าในปัจจุบันมีมากกว่า 425,000 ชุด และมียอดการดาวน์โหลดมากกว่า 15 พันล้านครั้งนับตั้งแต่เปิดตัวตลาดสำหรับเผยแพร่และจัดจำหน่ายซอฟต์แวร์ประยุกต์ (App Store) [1] ซึ่งต่อจากนี้ไปจะใช้คำว่าตลาดฯ

อย่างไรก็ตามในการที่ผู้ใช้จะได้มาซึ่งซอฟต์แวร์ประยุกต์ที่ตรงตามกับความต้องการ ผู้ใช้จำเป็นที่จะต้องค้นหาและซื้อซอฟต์แวร์ประยุกต์ผ่านทางตลาดฯ ที่ควบคุมการดำเนินงานโดยผู้พัฒนาระบบปฏิบัติการของอุปกรณ์เคลื่อนที่นั้นๆ ซึ่งระบบค้นคืนซอฟต์แวร์ประยุกต์ของตลาดฯ ปัจจุบันยังมีปัญหาประสิทธิภาพในการค้นคืนโดยมีสาเหตุมาจากการที่ระบบค้นคืนซอฟต์แวร์ประยุกต์นั้นใช้หลักการจัดเรียงรายการผลการค้นคืนตามความสำคัญโดยพิจารณาจากคำที่ปรากฏในรายละเอียดของซอฟต์แวร์ประยุกต์เป็นหลักซึ่งผู้พัฒนาซอฟต์แวร์ประยุกต์แต่ละรายจะเป็นผู้ให้ข้อมูลเกี่ยวกับซอฟต์แวร์ประยุกต์ที่พัฒนาขึ้นเอง แต่ข้อมูลรายละเอียดซอฟต์แวร์นี้อาจไม่ครบถ้วนหรือเพียงพอในการตัดสินใจได้เนื่องจากข้อมูลรายละเอียดซอฟต์แวร์ประยุกต์ที่บันทึกนั้นไม่ได้อธิบายลักษณะเฉพาะ (Characteristic) ที่แท้จริงหรือหน้าที่การทำงาน (Functionality) ของซอฟต์แวร์ประยุกต์ นอกจากนี้ผู้พัฒนาบางรายมีเจตนาจงใจที่จะบันทึกข้อมูลรายละเอียดคลาดเคลื่อนเพื่อหวังผลทางการตลาดโดยการใช้คำสำคัญ (Keyword) ที่เป็นรู้จักในขณะนั้นหรือทำการเผยแพร่ซอฟต์แวร์ประยุกต์ชุดเดียวกันแต่ดัดแปลงเปลี่ยนข้อมูลรายละเอียดให้แตกต่างกัน (Bulk Publishing) หรือที่เรียกกันว่าซอฟต์แวร์ประยุกต์ที่เป็นสแปม (Spam Application) เพื่อให้ระบบค้นคืนซอฟต์แวร์ประยุกต์มีโอกาสที่จะพบซอฟต์แวร์ประยุกต์ได้ง่ายขึ้น [2] โดยผลกระทบที่เกิดขึ้นนี้ส่งผลให้การจัดเรียงรายการซอฟต์แวร์ประยุกต์ที่ค้นคืนมาได้นั้นไม่ตรงตามความต้องการที่จริงของผู้สืบค้น

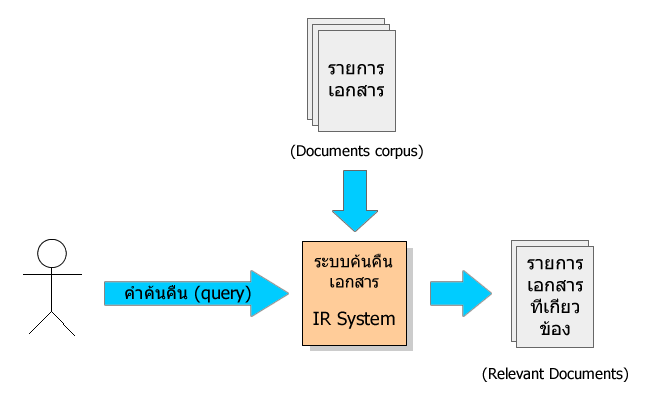
งานวิจัยนี้จึงมีจุดประสงค์เพื่อนำเสนอวิธีการปรับปรุงกระบวนการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ (Mobile Application Retrieval) ให้ตรงตามกับความต้องการความต้องการของผู้ใช้งานมากยิ่งขึ้น โดยนำอัลกอริทึมเพจแร็งค์ (PageRank Algorithm) ที่ใช้ในการวิเคราะห์ความสำคัญและจัดเรียงลำดับในการค้นคืนเว็บเพจมาประยุกต์ใช้กับการค้นคืนซอฟต์แวร์ประยุกต์ โดยข้อมูลที่จะนำมาใช้วิเคราะห์ความสำคัญของซอฟต์แวร์ประยุกต์นั้นจะพิจารณาจากข้อมูลรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ และข้อมูลรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นติดตั้งเป็นหลัก ซึ่งเป็นข้อมูลที่ผู้พัฒนาและผู้ใช้งานไม่สามารถดัดแปลงได้ อีกทั้งเป็นข้อมูลที่ได้จากการใช้งานจริงของผู้ใช้รายอื่นด้วย โดยตัวอย่างข้อมูลซอฟต์แวร์ประยุกต์ที่ใช้ในงานวิจัยนี้ ทางผู้วิจัยมีความตั้งใจว่าจะใช้ตัวอย่างข้อมูลจากแอนดรอยด์มาร์เก็ต (Android Market) ซึ่งเป็นหนึ่งในตลาดฯ ที่ใช้ระบบปฏิบัติการกูเกิลแอนดรอยด์ (Google Android) ที่มีผู้ใช้เป็นจำนวนมากและซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ที่จะทำการค้นคืนจะเป็นภาษาอังกฤษเท่านั้น

# ทฤษฏีที่เกี่ยวข้อง

## การจัดเก็บและค้นคืนข้อมูลสารสนเทศ (Information Storage And Retrieval)

เทคโนโลยีการค้นคืนข้อมูล [4] คือ เทคโนโลยีที่ว่าด้วยการจัดเก็บ ประมวลผล ค้นคืน และนำเสนอข้อมูลและเอกสารที่มีความสัมพันธ์กับสิ่งที่ผู้ใช้ต้องการค้นคืน (Relevant Documents) ซึ่งเทคโนโลยีการค้นคืนข้อมูลนี้เกิดในช่วงทศวรรษที่ 1960 โดยในช่วงแรกการค้นคืนข้อมูลยังจำกัดอยู่เพียงเอกสารจำนวนไม่มาก แต่จากการเกิดขึ้นของอินเตอร์เน็ตช่วงทศวรรษที่ 1990 ส่งผลให้ข้อมูลมีปริมาณเพิ่มขึ้นอย่างรวดเร็ว ความจำเป็นที่ตามมาคือความต้องการในการค้นหาเอกสาร ด้วยเหตุนี้เองจึงมีการศึกษาวิจัยเทคนิคและวิธีการที่ทำให้การค้นคืนข้อมูลมีประสิทธิภาพมากขึ้น [3]

โดยทั่วไปเทคนิคของการค้นคืนนั้นระบบจะทำการประมวลผลเอกสารที่อยู่ถูกเก็บอยู่ในคลังข้อมูล (Documents Corpus) เพื่อสกัดคำสำคัญ (Keyword) ที่ปรากฏในเอกสาร มาสร้างเป็นดัชนี (Indexes) เก็บไว้ในฐานข้อมูล จากนั้นเมื่อผู้ใช้ทำการค้นคืนข้อมูลด้วยคำค้นคืน (Query) ระบบจะทำการเปรียบเทียบคำค้นคืนกับรายการดัชนีที่บันทึกอยู่ในฐานข้อมูล จากนั้นจึงแสดงผลรายการเอกสารที่มีความสัมพันธ์นำเสนอแก่ผู้ใช้ ดังที่แสดงให้เห็นในรูปที่ 1

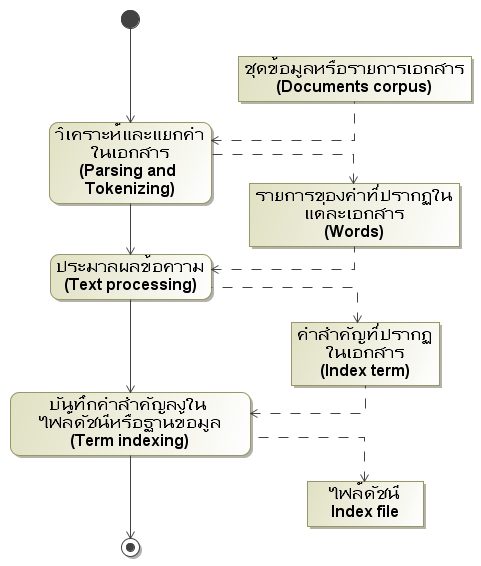


รูปที่ 1 การทำงานของระบบค้นคืนทั่วไป

กระบวนการทำงานของระบบการค้นคืน (Information Retrieval Engine) โดยหลักๆ จะถูกแบ่งเป็น 2 ขั้นตอนใหญ่ๆคือ

* + 1. ขั้นตอนการจัดเก็บข้อมูลและจัดทำดัชนีคำสำคัญ (Storage and indexing phase)

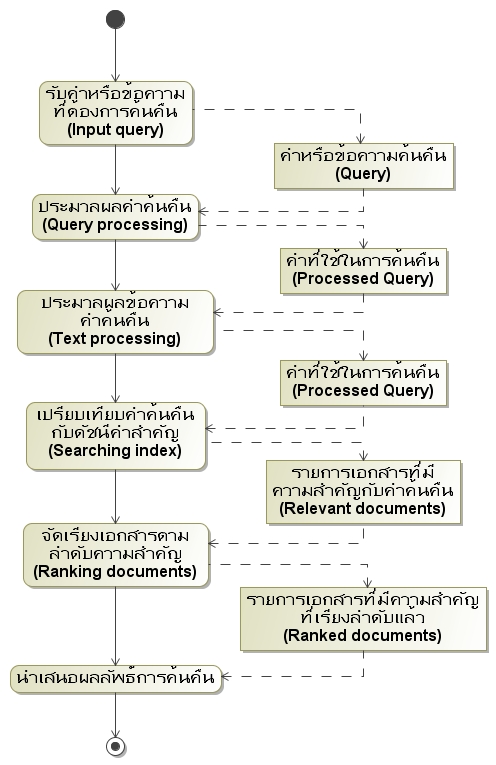
ในขั้นตอนนี้มีวัตถุประสงค์ในการจัดทำและบันทึกข้อมูลดัชนีคำสำคัญ (Index Terms) ซึ่งเป็นเสมือนตัวแทนที่ใช้ในการค้นคืนเอกสารในคลังเพื่อใช้ในการเปรียบเทียบกับคำค้นคืน (Query) โดยกระบวนการจะเป็นไปตามที่แสดงในรูปที่ 2 ซึ่งเริ่มจากการนำชุดข้อมูลหรือเอกสารที่ถูกจัดเก็บอยู่ในคลังมาทำการวิเคราะห์และแบ่งเป็นรายการของคำที่ปรากฏในแต่ละเอกสาร (Parsing and tokenizing) จากนั้นจึงนำคำที่ได้มาผ่านกระบวนการประมวลผลข้อความ (Text Processing) ด้วยการกรองคำที่ไม่มีความหมายออก (Stopword removal) เช่น คำว่า the, an, a เป็นต้น หรือแปลงคำสำคัญให้อยู่ในรูปรากศัพท์ (Word stemming) เช่นคำว่า computing หรือ computed จะถูกแปลงให้อยู่ในรูป compute สุดท้ายจึงนำข้อมูลที่ได้มาเก็บลงไฟล์ดัชนีคำสำคัญหรือฐานข้อมูล (Index file) โดยจะบันทึกคำสำคัญ รายการเอกสารที่คำสำคัญนั้นปรากฏอยู่ และจำนวนคำสำคัญที่ปรากฏอยู่ในเอกสาร (Term Frequency)



รูปที่ ขั้นตอนการจัดเก็บข้อมูลและจัดทำดัชนีคำสำคัญ

* + 1. ขั้นตอนการค้นคืนเอกสาร (Retrieval phase)

ในขั้นตอนนี้มีวัตถุประสงค์ในค้นคืนเอกสารที่ต้องการในคลังข้อมูล ซึ่งผลลัพธ์ที่ได้จะเป็นรายการเอกสารที่มีความสัมพันธ์กับสิ่งที่ผู้ใช้ต้องการค้นคืน โดยกระบวนการนี้จะเป็นไปตามที่แสดงในรูปที่ 3 ซึ่งเริ่มจากการที่ผู้ใช้ระบบทำการพิมพ์คำหรือข้อความที่ต้องการค้นหาเข้าสู่ระบบ จากนั้นระบบจะนำคำค้นหาที่รับมาไปทำการประมวลผลคำค้นหา (Query Processing) และประมวลผลข้อความ (Text Processing) ก่อนนำไปเปรียบเทียบตรวจสอบกับดัชนีคำสำคัญเพื่อให้ได้มาซึ่งรายการเอกสารที่มีคำที่ต้องการค้นหาปรากฏอยู่ หลังจากนั้นจึงนำเสนอผลลัพธ์แก่ผู้ใช้โดยในการนำเสนอรายการเอกสารนั้น ระบบจะทำการจัดเรียงตามลำดับความสำคัญของเอกสาร โดยพิจารณาจากจำนวนคำสำคัญที่ปรากฏในเอกสารหรือพิจารณาตามความต้องการของผู้ใช้



รูปที่ 3 ขั้นตอนการค้นคืนเอกสาร

## ตลาดซอฟต์แวร์ประยุกต์ (Application Market)

ตลาดซอฟต์แวร์ประยุกต์นั้นเป็นรูปแบบหนึ่งการกระจายดิจิทัล (Digital Distribution) ซึ่งมีวัตถุประสงค์ในการเผยแพร่และจัดจำหน่ายซอฟต์แวร์ประยุกต์บนแพลตฟอร์มต่างๆ โดยมีอินเตอร์เน็ตตัวกลางในการกระจายข้อมูลซอฟต์แวร์ประยุกต์ไปสู่ผู้ใช้เพื่อติดตั้งลงบนอุปกรณ์ของผู้ใช้งาน การกระจายข้อมูลในรูปแบบนี้มีข้อดีที่ไม่จำเป็นต้องใช้วัตถุตัวกลางในการบันทึกข้อมูลอย่างกระดาษ แผ่นซีดี หรือแผ่นดีวีดี ทำให้การกระจายข้อมูลนั้นเป็นไปได้ง่าย ลดค่าขนส่ง และมีค่าใช้จ่ายต่ำ

ในปัจจุบันนั้นมีตลาดซอฟต์แวร์ประยุกต์อยู่เป็นจำนวนมาก ในงานวิจัยนี้จึงเลือกที่จะศึกษาเฉพาะแอนดรอยด์มาร์เก็ต ซึ่งเป็นตลาดที่ใช้จัดจำหน่ายซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ที่ใช้ระบบปฏิบัติการกูเกิลแอนดรอยด์ (Google Android) [5] เนื่องจากแอนดรอยด์มาร์เก็ตให้ข้อมูลรายละเอียดซอฟต์แวร์ประยุกต์ที่สามารถนำมาใช้วิเคราะห์มากกว่าตลาดซอฟต์แวร์ประยุกต์อื่น เช่น ข้อมูลแสดงจำนวนผู้ติดตั้งซอฟต์แวร์ประยุกต์ หรือ รายการสถิติการจัดจำหน่ายซอฟต์แวร์ประยุกต์ในช่วงเวลาหนึ่งๆ เป็นต้น

ข้อมูลรายละเอียดของซอฟต์แวร์ประยุกต์ที่จำหน่ายในแอนดรอยด์มาร์เก็ต [5] ซึ่งผู้วิจัยสามารถนำมาวิเคราะห์ได้จะแสดงให้เห็นในตัวอย่างรูปที่ 4 ซึ่งมีรายละเอียดดังนี้

1. ชื่อของซอฟต์แวร์ประยุกต์
2. รายละเอียดของซอฟต์แวร์ประยุกต์
3. ตัวอย่างหน้าจอของซอฟต์แวร์ประยุกต์
4. คะแนนความนิยมของซอฟต์แวร์ประยุกต์ที่ผู้ติดตั้งซอฟต์แวร์เป็นผู้ให้คะแนน
5. ความคิดเห็นของผู้ที่ใช้งานซอฟต์แวร์
6. วันที่ล่าสุดที่ซอฟต์แวร์ประยุกต์ถูกปรับปรุง
7. เลขเวอร์ชันของซอฟต์แวร์ประยุกต์
8. เวอร์ชันของระบบปฏิบัติการขั้นต่ำที่ต้องการ
9. ประเภทของซอฟต์แวร์ประยุกต์ ตามการจัดประเภทซอฟต์แวร์ 22 ประเภท ที่กำหนดโดยแอนดรอยด์มาร์เก็ต
10. จำนวนผู้ใช้ที่ติดตั้งซอฟต์แวร์ประยุกต์
11. ข้อมูลกราฟสถิติการติดตั้งซอฟต์แวร์ประยุกต์ในช่วงเวลาหนึ่งๆ
12. ราคาของซอฟต์แวร์ประยุกต์
13. คะแนนเรตติ้งความรุนแรงของเนื้อหาในซอฟต์แวร์ประยุกต์
14. รายชื่อซอฟต์แวร์ประยุกต์ที่พัฒนาโดยนักพัฒนารายเดียวกัน
15. รายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ
16. รายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นติดตั้ง



รูปที่ ตัวอย่างหน้าเว็บเพจที่แสดงรายละเอียดซอฟต์แวร์ประยุกต์ในแอนดรอยด์มาร์เก็ต

## อัลกอริทึมเพจแร็งค์ (PageRank Algorithm)

อัลกอริทึมเพจแร็งค์นั้นเป็นอัลกอริทึมที่ใช้ในวิเคราะห์ความสำคัญ (Importance) ของเว็บเพจ [6] เพื่อใช้ในการจัดเรียงลำดับการค้นคืน (Ranking) เว็บเพจ โดยค่าเพจแร็งค์นั้นจะค่าตัวเลขที่ใช้แทนความสำคัญของแต่เว็บเพจ ซึ่งได้มาจากการวิเคราะห์จำนวนของเว็บเพจอื่นๆที่อ้างอิงถึงหน้าเว็บเพจ (Backlink) ที่ต้องการรู้ค่าเพจแร็งค์ แต่ปัญหาของอัลกอริทึมนี้คือไม่สามารถป้องกันการปลอมแปลงจำนวนลิงค์อ้างอิงได้ (Spam Link) ได้

ค่าเพจแร็งค์ของแต่ละเว็บเพจนั้นสามารถคำนวณได้จากสูตรสมการที่ 1 ซึ่งมีวิธีการการคำนวณ ดังต่อไปนี้

1. เมื่อเริ่มต้นกำหนดให้ค่าเพจแร็งค์ของแต่ของละเว็บเพจให้มีค่ารวมกันทั้งหมดเท่ากับ 1
2. คำนวณค่าเพจแร็งค์ของแต่ละเว็บเพจด้วยการใช้สูตรสมการ

(1)

โดยที่ คือ ค่าเพจแร็งค์ของเว็บเพจที่ถูกอ้างอิงถึง

คือ ค่าเพจแร็งค์ของเว็บเพจที่อ้างอิงมาหา

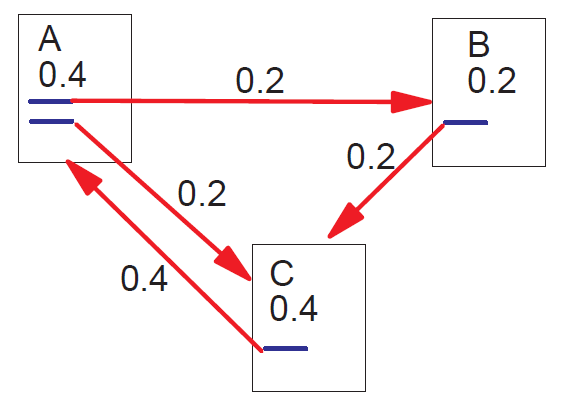
คือ จำนวนลิงค์อ้างอิงทั้งหมดที่ปรากฏอยู่ในเว็บเพจที่อ้างอิงถึงเว็บเพจที่ต้องการหา

ค่าเพจแร็งค์

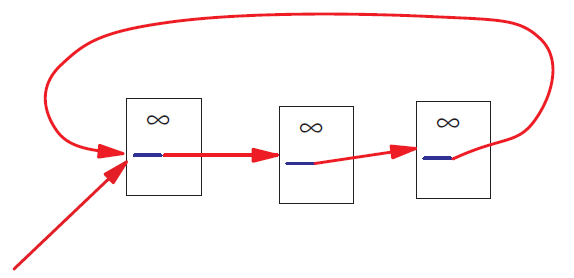
คือ เว็บเพจที่อ้างอิงถึงเว็บเพจที่ต้องการหาค่าเพจแร็งค์

1. ทำการคำนวณหาค่าเพจแร็งค์ให้ครบทุกเว็บเพจ
2. ตรวจสอบค่าเพจแร็งค์ใหม่ที่ได้ว่ามีความเสถียรในช่วงทศนิยมที่ต้องการหรือไม่ โดยคำนวณได้จากค่าเพจแร็งค์เดิมกับค่าเพจแร็งค์ใหม่ที่คำนวณได้นั้นมีค่าต่างกันไม่เกินที่กำหนดไว้
3. ถ้าค่าเพจแร็งค์ยังไม่เสถียร ให้ทำตามขั้นที่ 2 อีกครั้ง
4. เมื่อค่าเพจแร็งค์ใหม่ที่ได้ว่ามีความเสถียรในช่วงทศนิยมที่ต้องการแล้ว ให้หยุดการคำนวณและนำค่าที่ได้ไปทำการจัดลำดับผลการค้นคืนต่อไป

แต่จากการนำอัลกอริทึมเพจแร็งค์ไปใช้งานพบว่าการคำนวณเพจแร็งค์นั้นสามารถทำงานได้ถูกต้องในกรณีที่รูปแบบกราฟความสัมพันธ์ของเว็บเพจนั้นมีลักษณะการเชื่อมโยงทั่วถึงกันเท่านั้นดังที่แสดงในรูปที่ 5 ซึ่งในความเป็นจริงนั้นอาจเกิดกรณีที่เว็บเพจบางหน้านั้นไม่มีการอ้างอิงถึงเว็บเพจอื่นเลย ผลกระทบที่เกิดขึ้นคือทำให้เกิดปัญหาค่าอันดับรั่ว (Rank Leak) ส่งผลให้ผลรวมค่าเพจแร็งค์ของเว็บเพจที่คำนวณได้นั้นลดลงเรื่อยๆ ในแต่ละรอบการคำนวณ หรืออีกกรณีหนึ่งคือเว็บเพจนั้นมีการอ้างอิงไปยังอีกเว็บเพจและเว็บเพจนั้นมีการอ้างอิงกลับมายังเว็บเพจเดิม ผลกระทบที่เกิดขึ้นคือทำให้เกิดปัญหาค่าอันดับจม (Rank Sink) ส่งผลให้ค่าเพจแร็งค์ของเว็บเพจที่วนรอบอ้างอิงกันสูงกว่าเว็บเพจอื่น ดังที่แสดงในรูปที่ 6



รูปที่ กราฟความสัมพันธ์ของเว็บเพจนั้นมีลักษณะการเชื่อมโยงทั่วถึงกัน (Simplified PageRank) [6]



รูปที่ กราฟความสัมพันธ์ของเว็บเพจที่มีลักษณะความสัมพันธ์ที่วนรอบกันทำให้เกิดปัญหาค่าอันดับจม [6]

ปัญหาค่าอันดับรั่วและค่าอันดับจมนั้นสามารถแก้ไขได้สองวิธี วิธีแรกคือตัดเว็บเพจที่ก่อให้เกิดค่าอันดับรั่วออกจากการคำนวณ แต่วิธีนี้เป็นวิธีที่ไม่ได้รับความนิยมเนื่องจากจะทำให้ระบบไม่สามารถค้นคืนเว็บเพจที่ถูกตัดออก วิธีที่สองคือกำหนดค่าอัตราความเสื่อมสภาพหรือ Decay Factor () ที่มีค่าตั้งแต่ 0 ถึง 1 ให้กับแต่ละเว็บเพจในการคำนวณค่าเพจแร็งค์ โดยค่าอัตราความเสื่อมสภาพนั้นจะทำให้ค่าเพจแร็งค์ของทุกเว็บเพจที่คำนวณได้นั้นลู่เข้า (Convergence) เสมอ เสมือนสร้างเส้นการเชื่อมโยงเสมือนแบบทั่วถึงกัน (Virtual strongly connected hyperlinks)

การคำนวณค่าเพจแร็งค์ที่แก้ปัญหาค่าอันดับรั่วและค่าอันดับจมแล้วสามารถคำนวณได้ตามสูตรสมการที่ 2

(2)

โดยที่ คือ ค่าอัตราความเสื่อมสภาพที่ได้จากการทดลองไปเรื่อยๆ จนกว่าจะได้ค่าที่ดีที่สุดในการค้นคืน

คือ จำนวนเว็บเพจทั้งที่ใช้ในการคำนวณ

# งานวิจัยที่เกี่ยวข้อง

## AppJoy: Personalized Mobile Application Discovery

แอปจอยเป็นงานวิจัยและพัฒนาระบบเพื่อใช้ในการแนะนำ (Recommend) ซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ที่ใช้ระบบปฏิบัติการกูเกิลแอนดรอยด์ [8] โดยใช้เทคนิคตัวกรองร่วม (Collaborative Filtering) มาวิเคราะห์ข้อมูลการใช้งานและการติดตั้งซอฟต์แวร์ประยุกต์ของผู้ใช้มาเป็นตัวตัดสินใจ ซึ่งในงานวิจัยนี้ระบบจะทำการวิเคราะห์ข้อมูลและให้คะแนนความ “ชอบ” (Like) ของผู้ใช้งานซอฟต์แวร์ต่อซอฟต์แวร์ประยุกต์แต่ละตัวด้วยการพิจารณาค่า RFD ซึ่ง R ย่อมาจาก Recency หรือระยะเวลาระหว่างการใช้งานซอฟต์แวร์ในแต่ละครั้ง F ย่อมาจาก Frequency หรือจำนวนครั้งผู้ใช้เรียกใช้งานซอฟต์แวร์ในระยะเวลาหนึ่งๆ และ D ย่อมาจาก Duration หรือระยะเวลาที่ผู้ใช้เรียกใช้งานซอฟต์แวร์ในแต่ละครั้ง

การคำนวณค่า RFD นั้นสามารถจากเขียนออกมาได้ตามสมการที่ 3

 (3)

โดยที่  คือ ค่าความชอบของผู้ใช้งานซอฟต์แวร์ () ต่อซอฟต์แวร์ประยุกต์ ()

คือ ระยะเวลาที่ผู้ใช้เรียกใช้โปรแกรมนับตั้งแต่การใช้ครั้งสุดท้าย

คือ จำนวนครั้งผู้ใช้เรียกใช้งานซอฟต์แวร์ในระยะเวลาหนึ่งๆ

คือ ระยะเวลาที่ผู้ใช้เรียกใช้งานซอฟต์แวร์ในแต่ละครั้ง

คือ ค่าน้ำหนักของตัวแปร RFD แต่ละตัว โดยผู้ใช้งานระบบสามารถกำหนดได้

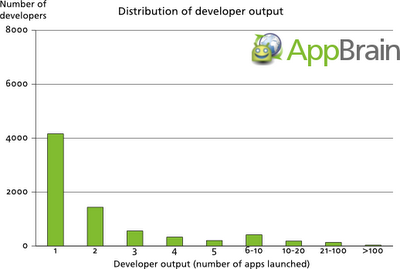
จากการคำนวณวิเคราะห์ข้อมูลการเปลี่ยนแปลงไปของค่า RFD ของผู้ใช้ซอฟต์แวร์แต่ละราย ระบบแอปจอยสามารถแนะนำซอฟต์แวร์ที่ผู้ใช้ส่วนใหญ่ชอบแก่ผู้ใช้รายอื่นที่มีความสนใจคล้ายคลึงกันได้ โดยในการประเมินผลการแนะนำซอฟต์แวร์ประยุกต์ของระบบแอปจอยนั้น ผู้วิจัยได้ประเมินผลประสิทธิภาพการแนะนำจากการดูข้อมูลการใช้งานระบบแอปจอยบนสภาพแวดล้อมจริง จากจำนวนผู้ใช้ 2,603 รายในช่วงเดือนมิถุนายน ปี 2010 ถึงเดือนกุมภาพันธ์ ปี 2011 พบว่าระบบแอปจอยสามารถแนะนำซอฟต์แวร์ประยุกต์ที่มีความถูกต้องมากกว่า 80% ให้กับผู้ใช้มากกว่า 80%

จากงานวิจัยนี้ผู้วิจัยแสดงให้เห็นว่า การนำข้อมูลการใช้งานซอฟต์แวร์มาใช้วิเคราะห์ความชอบของผู้ใช้ซอฟต์แวร์ประยุกต์แต่ละคนนั้นมีความเหมาะสมกว่าการใช้คะแนนความนิยมของซอฟต์แวร์ประยุกต์หรือจำนวนผู้ติดตั้งซอฟต์แวร์แต่ละตัว เพราะคะแนนความนิยมของซอฟต์แวร์ประยุกต์และจำนวนการติดตั้งซอฟต์แวร์นั้นเป็นการเก็บรวบรวมข้อมูลการใช้งานของซอฟต์แวร์ประยุกต์ทุกเวอร์ชัน ไม่ได้ทำการเก็บรวบรวมใหม่เมื่อมีการปรับปรุง ทำให้ค่าคะแนนที่ได้ไม่ได้แสดงถึงคุณภาพของซอฟต์แวร์ประยุกต์ที่จำหน่ายอยู่ในปัจจุบัน นอกจากนี้ผู้ใช้ส่วนใหญ่นั้นมักจะไม่ได้ให้คะแนนความนิยมของซอฟต์แวร์ประยุกต์ที่ตัวเองติดตั้ง

## AppBrain

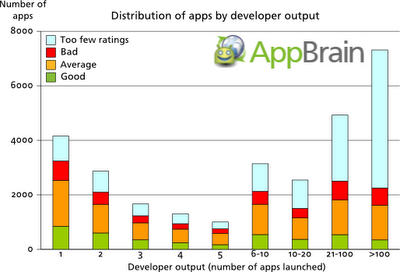
แอปเบรนเป็นระบบที่ใช้ในการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ที่ใช้ระบบปฏิบัติการกูเกิลแอนดรอยด์ ซึ่งถูกพัฒนาขึ้นโดยบริษัทเอกชนที่ชื่อว่า AppTornado GmbH [9] ความสามารถพิเศษของระบบค้นคืนนี้คือ ความสามารถในการกรองซอฟต์แวร์ประยุกต์ที่เป็นสแปมอัตโนมัติ (Automatic filtering of spam apps) โดยในการตรวจสอบซอฟต์แวร์ประยุกต์ที่เป็นสแปมนั้น ระบบแอปเบรนใช้การคำนวณเพื่อหาความน่าจะเป็นว่าผู้พัฒนารายใดมีโอกาสเป็นผู้เผยแพร่ซอฟต์ประยุกต์ที่เป็นสแปมหรือที่เรียกกันว่าสแปมเมอร์ (Spammer) ซึ่งพิจารณาได้จากพฤติกรรมที่ผิดปกติของผู้พัฒนาที่เผยแพร่ซอฟต์แวร์ที่ซอฟต์แวร์จำนวนมากเกินจนผิดปกติ

จากการผลวิจัยของบริษัทผู้พัฒนาระบบแอปเบรนระบุว่าในเดือนมีนาคมปี พ.ศ.2553 จำนวนซอฟต์แวร์ประยุกต์ที่ขายในแอนดรอยด์มาร์เก็ตของประเทศสหรัฐอเมริกามีทั้งหมด 28,936 ชุด ซึ่งพัฒนาโดยผู้พัฒนาจำนวน 7,429 ราย โดยเฉลี่ยผู้พัฒนาแต่ละรายเผยแพร่ซอฟต์แวร์ประยุกต์จำนวน 3.9 ชุด แต่ในรายชื่อผู้พัฒนานั้นพบว่ามีผู้พัฒนาจำนวน 32 รายที่เผยแพร่ซอฟต์แวร์ประยุกต์มากกว่า 100 ชุด ตามที่แสดงในรูปที่ 7



รูปที่ 7 กราฟแสดงอัตราส่วนจำนวนการเผยแพร่ซอฟต์แวร์ประยุกต์ในแอนดรอยด์มาร์เก็ตต่อจำนวนผู้พัฒนาทั้งหมดช่วงเวลามีนาคมปี พ.ศ. 2553 [9]

อาศัยสถิติข้อมูลการเผยแพร่ซอฟต์แวร์ประยุกต์ ผู้พัฒนาระบบแอปเบรนได้ทำการวิเคราะห์ข้อมูลว่าซอฟต์แวร์ประยุกต์ที่พัฒนาออกมานั้นมีคุณภาพหรือไม่ด้วยการพิจารณาจากคะแนนความนิยมที่ได้จากผู้ใช้งาน โดยผลที่ได้นั้นผู้พัฒนาระบบแอปเบรนได้ทำการแยกซอฟต์แวร์ประยุกต์ออกมาเป็น 4 ระดับได้แก่ ซอฟต์แวร์ประยุกต์ที่มีคุณภาพดี (Good) หรือได้คะแนนเฉลี่ยเกิน 4 คะแนนจากทั้งหมด 5 คะแนน ซอฟต์แวร์ประยุกต์ที่มีคุณภาพปานกลาง (Average) หรือได้คะแนนเฉลี่ยเกิน 3 คะแนนแต่ไม่ถึง 4 คะแนน ซอฟต์แวร์ประยุกต์ที่มีคุณภาพแย่ (Bad) หรือได้คะแนนเฉลี่ยต่ำกว่า 3 และซอฟต์แวร์ประยุกต์ที่ไม่สามารถประเมินคุณภาพได้เนื่องจากมีจำนวนผู้ใช้ให้คะแนนความนิยมน้อยเกินไป (Too Few Rating) ตามที่แสดงในรูปที่ 8



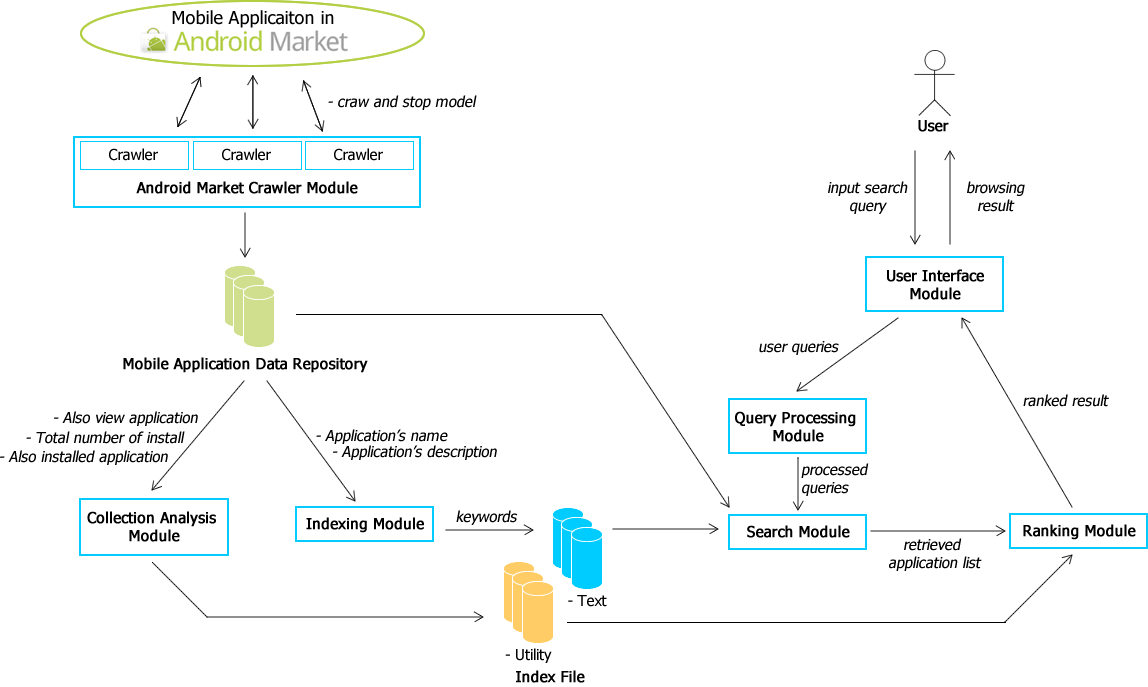
รูปที่ กราฟแสดงอัตราส่วนคะแนนความนิยมของซอฟต์แวร์ประยุกต์แต่ละชุดต่อจำนวนซอฟต์แวร์ที่พัฒนาโดยผู้พัฒนารายเดียวกัน [9]

จากผลการวิเคราะห์ข้อมูลที่ได้ ผู้พัฒนาระบบแอปเบรนได้ตั้งสมมุติฐานว่าผู้พัฒนาซอฟต์แวร์ที่เผยแพร่ซอฟต์แวร์ประยุกต์เป็นจำนวนมากแต่ซอฟต์แวร์ประยุกต์ส่วนใหญ่ที่พัฒนาขึ้นนั้นมีคุณภาพแย่ มีความเป็นไปได้ว่าจะ สแปมเมอร์สูง ผู้พัฒนาระบบแอปเบรนจึงสร้างตัวกรองซอฟต์แวร์ประยุกต์ที่คาดว่าจะเป็นสแปม โดยการทำคัดซอฟต์แวร์ประยุกต์ที่มาจากผู้พัฒนาที่มีความเป็นได้ว่าจะสแปมเมอร์สูงออกจากรายการค้นคืน ผลที่ได้จากการทดลองใช้ตัวกรองพบว่ามีผู้พัฒนาจำนวน 359 ราย ถูกระบุว่ามีความเป็นได้ว่าจะเป็นสแปมเมอร์ ที่เผยแพร่ซอฟต์แวร์ประยุกต์ที่ไม่มีคุณภาพมากกว่า 10,000 ชุดในแอนดรอยด์มาร์เก็ต โดยที่เหลือเป็นซอฟต์แวร์ประยุกต์ที่มีคุณภาพดี 14% และอีก 20% ซอฟต์แวร์ประยุกต์ที่ไม่สามารถประเมินคุณภาพได้

จากข้อสรุปของงานนี้ทำให้เราทราบได้ว่าแม้วิธีการกรองซอฟต์แวร์ประยุกต์ด้วยการวิเคราะห์ความเป็นไปได้ของผู้พัฒนาที่มีพฤติกรรมเป็นสแปมเมอร์จะทำให้ค้นคืนซอฟต์แวร์ประยุกต์จะมีประสิทธิภาพดีขึ้น แต่วิธีการนี้ยังมีปัญหาที่อาจวิเคราะห์ผู้พัฒนาผิดได้ เนื่องจากซอฟต์แวร์ที่มีคุณภาพต่ำไม่สามารถนิยามได้ว่าเป็นซอฟต์แวร์ที่มีลักษณะการทำงานเหมือนกัน อีกทั้งในบางกรณีการวิเคราะห์ไม่สามารถทำได้หากมีข้อมูลคะแนนความนิยมของผู้ใช้ไม่เพียงพอ

# แนวคิดและวิธีวิจัย

งานวิจัยนี้จะถูกแบ่งเป็น 4 ขั้นตอนหลักคือ 1) ขั้นตอนพัฒนาระบบเพื่อใช้สำหรับจัดเก็บข้อมูลซอฟต์แวร์ประยุกต์ 2) ขั้นตอนพัฒนาระบบสำหรับวิเคราะห์และค้นคืนข้อมูลซอฟต์แวร์ประยุกต์ 3) ขั้นตอนทดสอบและประเมินผลการค้นคืนซอฟต์แวร์ประยุกต์ 4) ขั้นตอนพัฒนาระบบต้นแบบสำหรับค้นคืนข้อมูลซอฟต์แวร์ประยุกต์ที่สามารถใช้งานจริงได้ โดยระบบที่พัฒนาขึ้นจะมีลักษณะการทำงานตามที่แสดงในรูปที่ 9



รูปที่ 9 ตัวอย่างระบบต้นแบบสำหรับค้นคืนข้อมูลซอฟต์แวร์ประยุกต์ที่จะพัฒนาขึ้น

## ขั้นตอนพัฒนาระบบเพื่อใช้สำหรับจัดเก็บข้อมูลซอฟต์แวร์ประยุกต์

ในขั้นตอนพัฒนานี้มีจุดประสงค์ในการศึกษาและพัฒนาวิธีการที่ใช้ในการขูดและจัดเก็บข้อมูลซอฟต์แวร์ประยุกต์ที่ได้จากแอนดรอยด์มาร์เก็ต โดยผลลัพธ์ที่ได้ในขั้นตอนนี้คือคลังข้อมูลซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ (Mobile Application Data Repository) และข้อมูลดัชนีคำสำคัญที่ใช้ในการเปรียบเทียบการค้นคืน (Index File) ซึ่งในขั้นตอนนี้มีการแบ่งการทำงานเป็นส่วนๆดังนี้

* + 1. ส่วนขูดข้อมูลจากแอนดรอยด์มาร์เก็ต (Android Market Crawler Module)

ในส่วนนี้ระบบจะทำการขูดข้อมูลรายละเอียดเกี่ยวกับซอฟต์แวร์ประยุกต์ที่อยู่ในแอนดรอยด์มาร์เก็ตเพื่อนำมาเก็บบันทึกและใช้วิเคราะห์การค้นคืน โดยข้อมูลที่จะขูดมาใช้งานเป็นหลักในงานวิจัยนี้ได้แก่

1. ชื่อของซอฟต์แวร์ประยุกต์
2. รายละเอียดของซอฟต์แวร์ประยุกต์
3. จำนวนผู้ใช้ที่ติดตั้งซอฟต์แวร์ประยุกต์
4. รายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ
5. รายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นติดตั้ง

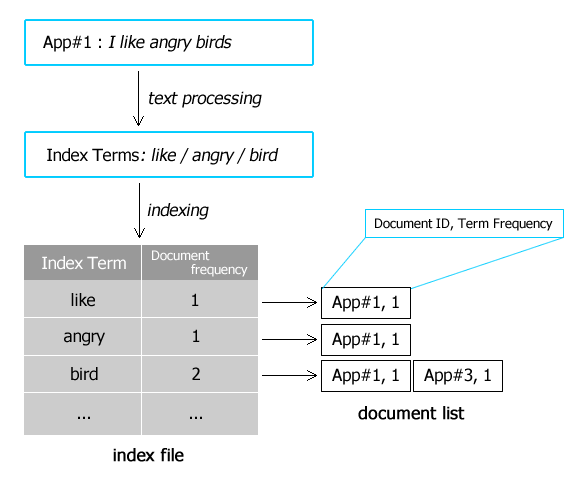
แต่เนื่องจากเนื่องจำนวนซอฟต์แวร์ประยุกต์ในตลาดฯนั้นมีจำนวนมาก ผู้วิจัยจึงเลือกแบบจำลองการขูดข้อมูล (Crawler Model) แบบ Crawl & Stop with Threshold มาใช้ โดยลิงค์ที่จะใช้ในการพิจารณาขูดข้อมูลต่อเนื่องจะ พิจารณาจากรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นติดตั้งและรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ ซึ่งหลังจากที่ได้ข้อมูลซอฟต์แวร์ประยุกต์มาแล้วข้อมูลทั้งหมดจะถูกเก็บอยู่ในฐานข้อมูลคลังข้อมูลซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่

* + 1. ส่วนสร้างดัชนีคำสำคัญ (Indexing Module)

หลังจากที่ได้ข้อมูลซอฟต์แวร์ประยุกต์มาแล้วนั้น ระบบจะนำข้อมูลที่ได้มาทำการสร้างข้อมูลดัชนีคำสำคัญ โดยในการสร้างข้อมูลดัชนีคำสำคัญจะใช้ข้อมูลชื่อและรายละเอียดของซอฟต์แวร์ประยุกต์มาผ่านกระบวนการประมวลผล เพื่อสกัดคำสำคัญที่ใช้สามารถแทนข้อมูลชื่อและรายละเอียดของซอฟต์แวร์ประยุกต์ในการเปรียบเทียบในการค้นคืน โดยขั้นตอนการวิเคราะห์และประมวลผลข้อความทำงานดังนี้

1. ตรวจการเรียงตัวของอักขระและตัดคำ (Parsing and Tokenizing)
2. กรองคำที่ไม่มีความหมายออก (Stopword Removal)
3. แปลงคำให้อยู่ในรากศัพท์ (Stemming and Plural Removal)

คำสำคัญที่ได้มานั้น ระบบจะนำบันทึกในในไฟล์ดัชนีคำสำคัญที่สร้างขึ้น ซึ่งไฟล์ดัชนีนี้จะอยู่ในรูปกลุ่มของคำต่างๆ พร้อมทั้งหมายเลขรหัสของซอฟต์แวร์ประยุกต์ที่คำเหล่านี้ปรากฏอยู่ (Inverted File) ตามที่แสดงให้เห็นดังรูปที่ 10



รูปที่ ตัวอย่างขั้นตอนการสกัดคำสำคัญจากชื่อซอฟต์แวร์ประยุกต์เพื่อมาเก็บบันทึกในไฟล์ดัชนี

* + 1. ส่วนวิเคราะห์ข้อมูลซอฟต์แวร์ประยุกต์ (Collection Analysis Module)

ในส่วนนี้จะเป็นการข้อมูลซอฟต์แวร์ประยุกต์ที่ได้มานั้น มาทำการสร้างเป็นดัชนีอรรถประโยชน์ (Utility index) โดยรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ รายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นติดตั้ง และจำนวนผู้ใช้ที่ติดตั้งซอฟต์แวร์ประยุกต์ จะถูกบันทึกเก็บไว้เพื่อนำไปใช้ในการวิเคราะห์จัดลำดับการค้นคืนต่อไป

## ขั้นตอนพัฒนาระบบสำหรับวิเคราะห์และค้นคืนข้อมูลซอฟต์แวร์ประยุกต์

ในขั้นตอนพัฒนานี้มีจุดประสงค์ในการศึกษาและพัฒนาวิธีการที่ใช้ในการค้นคืนข้อมูลซอฟต์แวร์ประยุกต์ที่ถูกเก็บอยู่ในคลังข้อมูลซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่และแสดงผลลัพธ์ที่ได้ให้ผู้ใช้ โดยผลลัพธ์ที่ได้ในขั้นตอนนี้คือรายการซอฟต์แวร์ประยุกต์ที่มีความสัมพันธ์กับคำค้นคืนโดยเรียงลำดับตามความสำคัญด้วยวิธีที่พัฒนาขึ้น ซึ่งในขั้นตอนนี้มีการแบ่งการทำงานเป็นส่วนๆดังนี้

* + 1. ส่วนประมวลผลคำค้นคืนจากผู้ใช้ (Query Processing Module)

ในส่วนนี้ระบบจะทำรับและประมวลคำค้นคืน (Query) ที่รับมาจากผู้ใช้ เพื่อทำการตรวจสอบสัญลักษณ์ในการเชื่อมคำเช่นเครื่องหมายตรรกะ “AND” “OR” หรือ “NOT” เป็นต้น เพื่อให้รู้ความต้องการของผู้ใช้ จากนั้นจึงส่งให้แก่ส่วนสืบค้นต่อไป

* + 1. ส่วนสืบค้นซอฟต์แวร์ประยุกต์ (Searching Module)

ในส่วนนี้ระบบจะทำการสืบค้นรายการซอฟต์แวร์ประยุกต์โดยใช้คำค้นคืนที่ได้รับมาจากผู้ใช้ โดยผลลัพธ์ที่ได้จากการสืบค้นคือรายการซอฟต์แวร์ประยุกต์ที่มีความสัมพันธ์กับคำค้นคืน ซึ่งมีขั้นตอนสืบค้นดังนี้

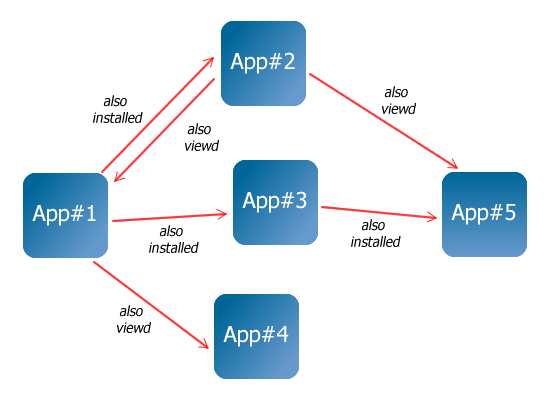
1. เปรียบเทียบคำค้นคืนกับคำในดัชนีคำสำคัญที่สร้างเก็บไว้
2. เมื่อพบคำค้นคืนที่ตรงกับคำสำคัญในดัชนี ให้ทำการดึงรายชื่อซอฟต์แวร์ประยุกต์ที่คำสำคัญนั้นปรากฏอยู่มารวบรวมไว้

รายการซอฟต์แวร์ประยุกต์ที่ได้มานั้นจะนำมาประมวลผลตามสัญลักษณ์ในการเชื่อมคำในขั้นตอนก่อนหน้า เพื่อพิจารณาว่าผู้ใช้ต้องการรายการผลลัพธ์การค้นคืนอย่างไร

* + 1. ส่วนจัดลำดับการผลลัพธ์การค้นคืน (Ranking Module)

ในส่วนนี้ระบบจะทำหน้าที่จัดเรียงลำดับผลการค้นคืนซอฟต์แวร์ประยุกต์ โดยจะเรียงลำดับรายการซอฟต์แวร์ประยุกต์ที่มีความสัมพันธ์กับคำค้นคืนที่ผู้ใช้ใส่มากที่สุด (High Relevance) ซึ่งในส่วนนี้จะเป็นในส่วนหลักของงานวิจัยของงานวิจัยนี้ โดยวิธีที่จะนำมาใช้คำนวณในการจัดลำดับการค้นคืนจะอ้างอิงวิธีคำนวณจากอัลกอริทึมเพจแร็งค์ ที่ใช้ในการวิเคราะห์ความสำคัญของเว็บเพจ แต่ในที่นี้จะเป็นการวิเคราะห์เพื่อจัดลำดับความสำคัญซอฟต์แวร์ประยุกต์แทน

ปัญหาหนึ่งในการประยุกต์ใช้อัลกอริทึมเพจแร็งค์ มาใช้ในการวิเคราะห์ความสำคัญของซอฟต์แวร์ประยุกต์นั้นคือ ซอฟต์แวร์ประยุกต์นั้นไม่มีสิ่งที่ใช้เชื่อมความสัมพันธ์เหมือนกับหน้าเว็บเพจที่มีลิงค์ซึ่งใช้การอ้างอิงหากัน ดังนั้นในการแก้ปัญหานี้ผู้วิจัยจึงจะนำข้อมูลรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นใช้ และรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจ มาใช้เชื่อมความสัมพันธ์ระหว่างซอฟต์แวร์ประยุกต์แทนดังที่แสดงในรูปที่ 11



รูปที่ ตัวอย่างกราฟความสัมพันธ์ของซอฟต์แวร์ประยุกต์

นอกจากนี้อีกเหตุผลที่ทำให้ควรเลือกรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นใช้และรายชื่อซอฟต์แวร์ประยุกต์ที่ผู้ใช้รายอื่นสนใจมาใช้ในการวิจัยนี้ คือ ข้อมูลรายชื่อซอฟต์แวร์ประยุกต์นั้นเป็นข้อมูลที่เกิดขึ้นจากการใช้งานจริงของผู้ใช้ซอฟต์แวร์ไม่สามารถดัดแปลงได้ ซึ่งแตกต่างจากลิงค์อ้างอิงที่อยู่ในเว็บเพจที่ใช้มาวิเคราะห์ในอัลกอริทึมเพจแร็งค์ ซึ่งอาจเป็นการปลอมแปลงจำนวนลิงค์อ้างอิงได้

ในการคำนวณหาค่าความสำคัญของซอฟต์แวร์ประยุกต์แต่ละชุดด้วยอัลกอริทึมเพจแร็งค์นั้น สามารถคำนวณได้ด้วยวิธีการดังต่อไปนี้

1. เมื่อเริ่มต้นกำหนดให้ค่าความสำคัญให้แต่ละซอฟต์แวร์ประยุกต์โดยให้มีค่ารวมกันทั้งหมดเท่ากับ 1
2. กำหนดค่าอัตราค่าเสื่อมสภาพ () ที่จะใช้ในการคำนวณ
3. คำนวณค่าความสำคัญของแต่ละซอฟต์แวร์ประยุกต์ด้วยการใช้สูตรสมการ

(4)

โดยที่ คือ ค่าความสำคัญของซอฟต์แวร์ประยุกต์ที่ถูกอ้างอิงถึง

คือ ค่าความสำคัญของซอฟต์แวร์ประยุกต์ที่อ้างอิงมาหา

คือ จำนวนซอฟต์แวร์ประยุกต์ทั้งหมดที่ผู้ใช้รายอื่นสนใจหรือติดตั้ง ของซอฟต์แวร์ประยุกต์ที่

อ้างอิงถึงซอฟต์แวร์ประยุกต์ที่ต้องการคำนวณหาค่าความสำคัญ

คือ ซอฟต์แวร์ประยุกต์ที่อ้างอิงถึงซอฟต์แวร์ประยุกต์ที่ต้องการหาค่าความสำคัญ

คือ ค่าอัตราความเสื่อมสภาพที่ได้จากการทดลองไปเรื่อยๆ จนกว่าจะได้ค่าที่ดีที่สุดในการ

ค้นคืน

คือ จำนวนซอฟต์แวร์ประยุกต์ทั้งที่ใช้ในการคำนวณ

1. ทำการคำนวณ ค่าความสำคัญของซอฟต์แวร์ประยุกต์ให้ครบทุกตัว
2. ตรวจสอบค่าความสำคัญใหม่ที่คำนวณได้ว่ามีความเสถียรในช่วงทศนิยมที่ต้องการหรือไม่ โดยคำนวณได้จากค่าความสำคัญเดิมกับค่าความสำคัญใหม่ที่คำนวณได้นั้นมีค่าต่างกันไม่เกินที่กำหนดไว้
3. ถ้าค่าความสำคัญของซอฟต์แวร์ประยุกต์ที่คำนวณมาได้ยังไม่เสถียร ให้ทำตามขั้นที่ 3 อีกครั้ง
4. เมื่อค่าความสำคัญของซอฟต์แวร์ประยุกต์ใหม่ที่ได้ว่ามีความเสถียรในช่วงทศนิยมที่ต้องการแล้ว ให้หยุดการคำนวณและนำค่าที่ได้ไปทำการจัดลำดับผลการค้นคืนและนำเสนอต่อผู้ใช้ต่อไป
   * 1. ส่วนเชื่อมต่อระหว่างผู้ใช้และระบบ (User Interface Module)

ในส่วนนี้ระบบจะส่วนต่อประสานระหว่างผู้ใช้กับระบบโดยผู้ใช้จะเป็นผู้กรอกคำค้นคืนให้ระบบไปทำการประมวลผลและค้นคืน จากนั้นระบบจึงนำรายการซอฟต์แวร์ประยุกต์ที่มีความสัมพันธ์กับค้นคืนที่ถูกจัดเรียงแล้วแสดงให้ผู้ใช้ต่อไป

## ขั้นตอนทดสอบและประเมินผลการค้นคืนซอฟต์แวร์ประยุกต์

ในขั้นตอนนี้มีจุดประสงค์ในการทดสอบและวัดผลการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ว่าตรงตามความต้องการของผู้ใช้หรือไม่ โดยทางผู้วิจัยได้กำหนดวิธีการประเมินผลโดยยึดแนวทางที่ใช้ประเมินอัลกอริทึมฮิต (HITS – Hypertext-Induced Topic Search) [10] ที่ใช้วิเคราะห์ความสำคัญของเว็บเพจเช่นเดียวอัลกอริทึมเพจแร็งค์ ซึ่งในการประเมินนั้นจะเป็นการให้ผู้ใช้งานจำนวนหนึ่งมาให้คะแนนผลการค้นคืนของระบบที่พัฒนาอัลกอริทึมฮิตเทียบกับผลการค้นคืนของระบบอื่นที่ใช้คำค้นคืนเดียวกัน จากนั้นจึงนำผลคะแนนที่ได้มาประเมินผลและวิเคราะห์อีกทีหนึ่ง

สำหรับงานวิจัยนี้เบื้องต้นผู้วิจัยมีการกำหนดหลักการทดสอบในการประเมินไว้ดังต่อไปนี้

1. ผู้ทดสอบจะต้องเป็นผู้เคยใช้หรือใช้งานโทรศัพท์อุปกรณ์เคลื่อนที่แบบสมาร์ทโฟนหรืออุปกรณ์แท็บเล็ตคอมพิวเตอร์
2. คำค้นคืนที่ใช้ในการทดสอบนั้นจะสร้างโดยพิจารณาแยกตามประเภทซอฟต์แวร์ประยุกต์ 22 ประเภทที่กำหนดขึ้นโดยแอนดรอยด์มาร์เก็ต
3. คำค้นคืนที่ใช้ในการทดสอบนั้นจะมีทั้งคำที่เป็นชื่อเฉพาะ คำทั่วไปที่ไม่มีความสำคัญหรือเป็นที่รู้จัก และคำทั่วไปที่มีความสำคัญหรือเป็นที่รู้จัก
4. คำค้นคืนที่ใช้ในการทดสอบการค้นคืนของผู้ใช้แต่ละคนนั้นจะเป็นข้อมูลชุดเดียวกัน
5. ผู้ทดสอบจะให้คะแนนโดยการพิจารณาจากรายการซอฟต์แวร์ประยุกต์ 10 อันดับแรกที่ได้จากการค้นคืน
6. ผู้ทดสอบเป็นผู้ให้คะแนนว่าซอฟต์แวร์ประยุกต์แต่ละรายการที่ได้จากการค้นคืนนั้นตรงตามความต้องการหรือไม่ด้วยการให้คะแนน 4 ระดับ คือ แย่ (Bad) พอใช้ (Fair) ดี (Good) หรือยอดเยี่ยม (Fantastic)
7. ในการทดสอบจะเปรียบเทียบผลลัพธ์การค้นคืนกับระบบค้นคืน ดังต่อนี้
   1. ผลลัพธ์การค้นคืนจากระบบที่พัฒนาขึ้น
   2. ผลลัพธ์การค้นคืนจากแอนดรอยด์มาร์เก็ต โดยเรียงตามลำดับตามการคำนวณค่าความนิยมของแอนดรอยด์มาร์เก็ต
   3. ผลลัพธ์การค้นคืนจากแอนดรอยด์มาร์เก็ต โดยเรียงตามลำดับตามค่าคะแนนความนิยมที่ได้รับจากผู้ใช้ของซอฟต์แวร์ประยุกต์
   4. ผลลัพธ์การค้นคืนจากแอนดรอยด์มาร์เก็ต โดยเรียงตามลำดับตามจำนวนผู้ใช้ซอฟต์แวร์ประยุกต์

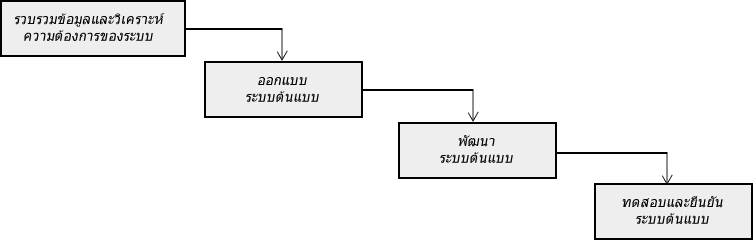
ผลลัพธ์ที่ได้จากการประเมินการผลค้นคืนซอฟต์แวร์ประยุกต์จะนำมาคำนวณหาเฉลี่ย () และค่าเบี่ยงเบนมาตรฐาน () เพื่อนำไปใช้วิเคราะห์ทางสถิติเพื่อเปรียบเทียบคุณภาพผลลัพธ์การค้นคืนที่ได้ระหว่างระบบค้นคืนที่พัฒนาขึ้นและระบบค้นคืนของแอนดรอยด์มาร์เก็ตต่อไป

## ขั้นตอนพัฒนาระบบต้นแบบสำหรับค้นคืนข้อมูลซอฟต์แวร์ประยุกต์ที่สามารถใช้งานจริงได้

ในขั้นตอนพัฒนานี้มีจุดประสงค์ในการพัฒนาระบบต้นแบบที่สามารถใช้ค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ได้ โดยในการพัฒนาผู้วิจัยได้มีการออกแบบและวางแผนการพัฒนาระบบเบื้องต้นไว้ดังนี้

* + 1. แบบจำลองที่ใช้ในการพัฒนาระบบ

ในการพัฒนาระบบต้นแบบนั้น ผู้วิจัยจะทำการพัฒนาระบบด้วยการใช้แบบจำลองการพัฒนาซอฟต์แวร์แบบวอเตอร์ฟอล (Water Model) ตามที่แสดงในแผนภาพที่ 12



รูปที่ 12 แผนภาพการพัฒนาซอฟต์แวร์ด้วยแบบจำลองวอเตอร์ฟอล

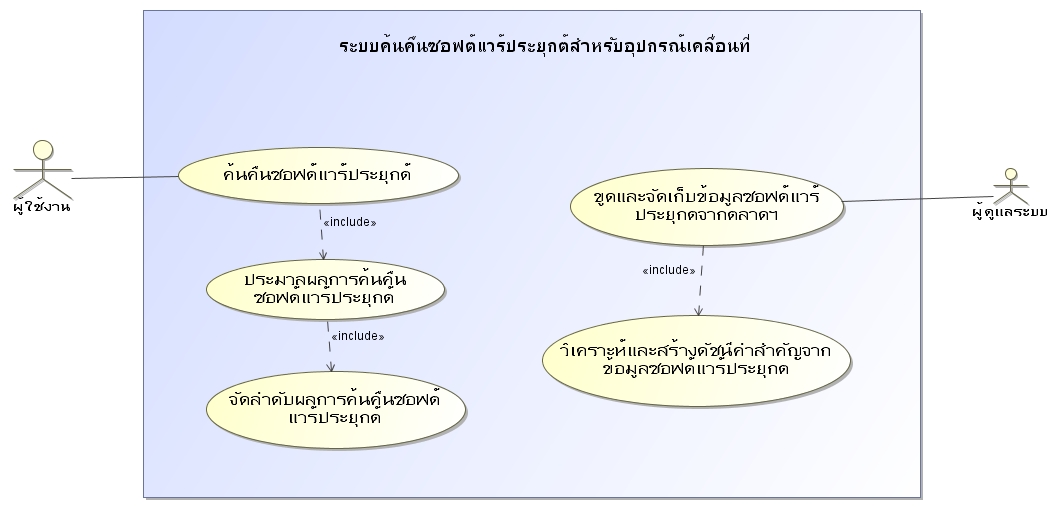
* + 1. การออกแบบสถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบค้นคืนที่จะพัฒนาขึ้นนั้น จะมีลักษณะสถาปัตยกรรมแบบเอ็มวีซี (MVC) และเชื่อมต่อกับผู้ใช้ผ่านโพรโทคอลทีซีพีไอพี (TCP/IP) เนื่องจากสถาปัตยกรรมแบบเอ็มวีซีนั้นจะแยกระบบออกเป็นส่วนๆ ทำให้การการพัฒนา การทดสอบ และการดูแลรักษาระบบแยกออกจากกัน ทำให้ง่ายแก่การเปลี่ยนแปลงและแก้ไขในภายหลัง

สถาปัตยกรรมของระบบค้นคืนสามารถแบ่งออกได้เป็น 3 ส่วน ได้แก่

1. ส่วนโมเดล (Model) เป็นส่วนที่ใช้ในการเข้าถึงและใช้งานคลังข้อมูลซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ และดัชนีคำสำคัญ
2. ส่วนวิว (View) เป็นส่วนที่นำข้อมูลที่ได้จากส่วนโมเดลมาแสดงผลให้ผู้ใช้งานระบบ
3. ส่วนคอนโทรลเลอร์ (Controller) เป็นส่วนรับข้อมูลคำค้นคืนจากผู้ใช้และทำการประมวลผลการค้นคืน โดยการให้ส่วนโมเดลค้นคืนซอฟต์แวร์ประยุกต์ที่สัมพันธ์กับคำค้นคืนก่อน แล้วจึงนำผลการค้นค้นที่ได้มาจัดลำดับตามความสำคัญ
   * 1. การออกแบบหน้าที่การทำงานของระบบ

ในการออกแบบหน้าที่การทำงานของระบบค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ ได้ดำเนินการออกแบบโดยเริ่มจากการศึกษาระบบการค้นคืนซอฟต์แวร์ประยุกต์ที่มีอยู่ในปัจจุบัน แล้วมาวิเคราะห์ความต้องการของระบบ ซึ่งแสดงเป็นแผนภาพยูสเคสดังรูปที่ 13



รูปที่ 13 แผนภาพยูสเคสของระบบค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่

# วัตถุประสงค์ของงานวิจัย

นำเสนอวิธีปรับปรุงและระบบต้นแบบที่ใช้ในการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ เพื่อให้การค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ตรงตามความต้องการของผู้ใช้งานมากยิ่งขึ้น

# ขอบเขตงานวิจัย

* 1. พัฒนาระบบต้นแบบที่ใช้ค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่จัดจำหน่ายและเผยแพร่ในตลาดซอฟต์แวร์ประยุกต์
  2. ประเมินความถูกต้องของระบบต้นแบบที่พัฒนาขึ้น ด้วยการเปรียบเทียบกับผลการค้นคืนจากแหล่งข้อมูลต้นทาง
  3. การค้นคืนซอฟต์แวร์ประยุกต์จะทำการค้นคืนซอฟต์แวร์ประยุกต์ที่มีรายละเอียดเป็นภาษาอังกฤษเท่านั้น

# 7. ขั้นตอนวิจัย

* 1. ศึกษาและทำความเข้าใจกระบวนการเผยแพร่และจัดจำหน่ายซอฟต์แวร์ประยุกต์ของตลาดซอฟต์แวร์ประยุกต์
  2. ศึกษาและวิเคราะห์ประเด็นปัญหาที่ต้องพิจารณาในการสร้างระบบค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่และงานวิจัยที่เกี่ยวข้อง
  3. ศึกษาและทำความเข้าใจวิธีการพัฒนาระบบค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่และเทคโนโลยีที่เกี่ยวข้อง
  4. ศึกษาและทดลองหาวิธีขูดข้อมูลที่ต้องการจากตลาดซอฟต์แวร์ประยุกต์ด้วยการใช้หลักการการค้นคืนข้อมูล
  5. ออกแบบและกำหนดลักษณะข้อมูลนำเข้าที่สามารถนำมาวิเคราะห์เพื่อใช้ในการค้นคืน
  6. เลือกเครื่องมือและพัฒนาระบบตามที่ได้ออกแบบไว้
  7. ทดสอบและประเมินผลวิธีวิจัยและระบบที่พัฒนาขึ้นเพื่อสนับสนุนแนวคิดในงานวิจัยนี้
  8. จัดทำบทความวิชาการ
  9. สรุปผลแนวทางการวิจัยและข้อเสนอแนะและจัดทำเล่มโครงงาน

# 8. ประโยชน์ที่คาดว่าจะได้รับ

* 1. ได้ระบบต้นแบบที่ใช้ในการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่
  2. ได้วิธีการค้นคืนซอฟต์แวร์ประยุกต์สำหรับอุปกรณ์เคลื่อนที่ที่สามารถค้นคืนข้อมูลได้ตรงตามความต้องการของผู้ใช้งานมากยิ่งขึ้น
  3. วิธีการที่พัฒนาขึ้นสามารถลดปัญหาการค้นคืนซอฟต์แวร์ประยุกต์ที่มีลักษณะเป็นซอฟต์แวร์ประยุกต์ที่เป็น สแปมได้

# 9. รายการอ้างอิง

[1] Apple Press Info, “Apple’s App Store Downloads Top 15 Billion” – July 7, 2011. http://www.apple.com/pr/library/2011/07/07Apples-App-Store-Downloads-Top-15-Billion.html

[2] The Windows Phone Developer Blog, “How Marketplace Manages Bulk App Publishing”. http://windowsteamblog.com/windows\_phone/b/wpdev/archive/2011/06/02/how-marketplace-manages-bulk-app-publishing.aspx

[3] ศูนย์บริการความรู้ทางวิทยาศาสตร์และเทคโนโลยี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ, “คลังศัพท์ไทย Thai Word Repository - Information storage and retrieval system”. http://www.thaiglossary.org/node/10434

[4] R. Baeza-Yates and B. Ribeiro-Neto, “Modern Information Retrieval,” ACM Press, Addison Wesley, 1999.

[5] Android Developers, http://developer.android.com/.

[6] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, and Sriram Raghavan. 2001. “Searching the Web” ACM Trans. Internet Technol. 1, 1 (August 2001), 2-43.

[7] Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) “The PageRank Citation Ranking: Bringing Order to the Web”. Proceedings of the 7th International World Wide Web Conference Brisbane, Australia: (1998), p. 161-172.

[8] Bo Yan and Guanling Chen. 2011. “AppJoy: personalized mobile application discovery”. InProceedings of the 9th international conference on Mobile systems, applications, and services (MobiSys'11). ACM, New York, NY, USA, 113-126.

[9] AppBrain. http://www.appbrain.com/.

[10] Jon M. Kleinberg. 1999. “Authoritative sources in a hyperlinked environment”. J. ACM 46, 5 (September 1999), 604-632.

[11] A. M. Davis, H. Bersoff, and E. R. Comer. 1988. A Strategy for Comparing Alternative Software Development Life Cycle Models. IEEE Trans. Softw. Eng. 14, 10 (October 1988), 1453-1461.