PULINET Journal

การวิเคราะห์คำค้น ในรูปแบบ Word cloud เพื่อสนับสนุนงานบริการสารสนเทศของห้องสมุด

Keywords Analysis in Word Cloud for Supporting Library Information Services

รตานา ยามาเจริญ
สำนักทรัพยากรการเรียนรู้คุณหญิงหลง อรรถกระวีสุนทร มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
e-mail: ratana.ya@psu.ac.th

บทคัดย่อ

งานวิจัยนี้นำเสนอวิธีการวิเคราะห์คำค้นภาษาอังกฤษที่ผู้ใช้บริการใช้ค้นหาข้อมูลบน ระบบโอแพคของสำนักทรัพยากรการเรียนรู้คุณหญิงหลง อรรถกระวีสุนทร ระหว่างเดือนมกราคม ถึง เดือนมีนาคม 2561 โดยการนำเทคนิคการทำเหมืองข้อความ (Text mining) ในรูปแบบ Word cloud มาใช้ในการวิเคราะห์ความถี่ของคำค้น (Keywords) และแสดงผลในรูปแบบ Visualization ขั้นตอนการดำเนินการประกอบด้วย 3 ขั้นตอนหลัก ดังนี้ 1) ก่อนดำเนินการ ประกอบด้วย การเก็บรวบรวมข้อมูล การเลือกข้อมูล และการทำความสะอาดข้อมูล 2) กำลังดำเนินการ ประกอบด้วย การวิเคราะห์ข้อมูล (Text Analysis) โดยใช้โปรแกรม R และ 3) หลังดำเนินการ ประกอบด้วย การแสดงผลการวิเคราะห์และสรุปข้อมูลคำค้น ในรูปแบบของ Word cloud และความถี่ของคำ ผลการวิจัยในครั้งนี้จะมีประโยชน์ทั้งผู้ใช้บริการและเจ้าหน้าที่ห้องสมุด ทำให้ทราบถึงแนวโน้มความต้องการและความสนใจของผู้ใช้บริการในการค้นหาข้อมูลทรัพยากรสารสนเทศของห้องสมุด ช่วยสนับสนุนการวางแผนการจัดกิจกรรมและการจัดหาทรัพยากรสารสนเทศให้สอดคล้องกับความต้องการและความสนใจของผู้ใช้บริการ และสามารถนำผลการวิจัยในครั้งนี้ เป็นเครื่องมือและแนวทางในการช่วยพัฒนาห้องสมุดเชิงรุกในอนาคตต่อไป

คำสำคัญ:

คีย์เวิร์ด, เวิร์ดคลาวด์, เหมืองข้อความ

Abstract

The purpose of this study was to analyze English keywords which users searches from OPAC of Khunying Long Athakravisunthorn Learning Resources Center during January to March 2018 by using text mining techniques in Word Cloud. The data was analyzing the frequency of keywords and display in the Visualization. The process consists of three main steps: 1) Before Pre-processing step consists of data collection,

data selection and data cleaning 2) Processing step consists of a text Analysis using the R program, and 3) Post-processing step consists of displaying, analyzing, and summarizing word data in the form of Word Cloud and word frequency.

The results of this Research will be useful to both users and library staff. Know the trends, needs and interests of users to find library information resources. Supports planning, library activities and acquiring information resources to meet the needs and interests of users. The results of this Research can be applied to be the tools and guidelines to develop libraries in the future.

Keyword:

Keyword, Word Cloud, Text Mining

บทนำ

ปัจจุบันปฏิเสธไม่ได้เลยว่า "ข้อมูล" เป็นสิ่งสำคัญมากในทุก ๆ วงการไม่ว่าจะเป็นทางด้านธุรกิจ ทางการแพทย์ และทางการศึกษา ถือเป็นทรัพยากรสารสนเทศที่มีคุณค่ามหาศาล อีกทั้งยังมีประโยชน์ต่อการสร้างองค์ความรู้ใหม่ ทุก ๆ หน่วยงานจึงเก็บรวบรวมข้อมูลและสร้างเป็นฐานข้อมูลไว้ โดยข้อมูลเหล่านี้สามารถนำไปวิเคราะห์ สังเคราะห์ให้เป็นสารสนเทศที่ใช้ในการตัดสินใจ หรือวางแผนในเรื่องต่าง ๆ เช่น การจัดซื้อ การจัดบริการ การแนะนำสินค้า เป็นต้น

สำนักทรัพยากรการเรียนรู้คุณหญิงหลง อรรถกระวีสุนทร ในฐานะเป็นแหล่งศึกษาค้นคว้าข้อมูล สำหรับอาจารย์ บุคลากร นักศึกษาและบุคคลทั่วไป ได้มีการเผยแพร่ความรู้ มีข้อมูลที่ได้จัดเก็บและรวบรวมข้อมูลไว้เป็นระบบจำนวนมาก อาทิเช่น ข้อมูลผู้เข้าใช้บริการห้องสมุด ข้อมูลการยืม-คืนทรัพยากรสารสนเทศ ข้อมูลการค้นหาทรัพยากรสารสนเทศ ข้อมูลการใช้อุปกรณ์หรือ ห้องศึกษาค้นคว้าด้วยตัวเอง ข้อมูลการเข้าถึงทรัพยากรสารสนเทศอิเล็กทรอนิกส์ต่าง ๆ โดยเฉพาะในการเข้าถึงทรัพยากรสารสนเทศของห้องสมุด ผู้ใช้บริการจำเป็นจะต้องบอกความต้องการของตัวเอง ผ่านการใช้คำค้นหรือที่เรียกว่า คีย์เวิร์ด เพื่อให้ทราบว่าผู้ใช้บริการคนนั้น ต้องการทรัพยากรสารสนเทศเรื่องใด คำค้นที่ผู้ใช้บริการใช้ เป็นเหมือนตัวแทนของทรัพยากรสารสนเทศที่ทำให้ผู้ใช้บริการสามารถเข้าถึงทรัพยากรสารสนเทศได้ตรงกับความต้องการมากที่สุด ในแต่ละวัน แต่ละช่วงเวลา ผู้ใช้บริการมีการค้นข้อมูลผ่านการใช้คำค้นจำนวนมาก บนระบบสืบค้นทรัพยากรสารสนเทศของห้องสมุด ทำให้มีคำค้นจำนวนมหาศาลถูกจัดเก็บรวบรวมไว้ในฐานข้อมูล ทั้งคำค้นที่เหมือนและแตกต่างกันจากผู้ใช้บริการ แต่คำค้นที่ได้ถูกเก็บรวบรวมไว้ก็ไม่ได้ถูกนำมาใช้ให้เกิดประโยชน์ ทำให้ ไม่ทราบว่าคำค้นใดบ้างที่มีการใช้บ่อย เนื้อหาประเภทใดที่ผู้ใช้ให้ความสนใจหรือต้องการในแต่ละช่วงเวลา ทำให้ไม่สามารถเข้าถึงหรือทราบความต้องการของผู้ใช้บริการได้

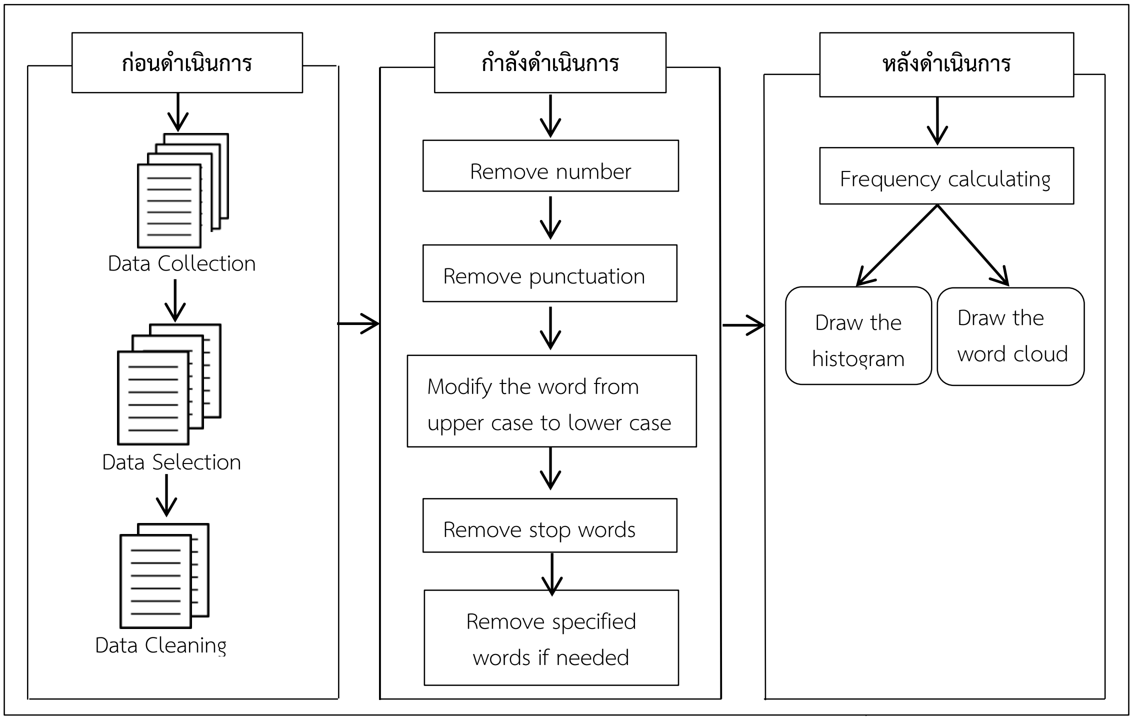
ข้อมูลคำค้น เป็นข้อมูลจริงที่มีประโยชน์ มีปริมาณการเพิ่มขึ้นและแตกต่างกันในทุกขณะ ผู้วิจัยจึงมีแนวคิดในการนำข้อมูลคำค้นมาวิเคราะห์และสรุปให้เกิดประโยชน์ เพื่อให้ทราบถึงแนวโน้มความต้องการและตระหนักถึงความสนใจของผู้ใช้บริการต่อทรัพยากรสารสนเทศที่มีในห้องสมุด ซึ่งจะเป็นประโยชน์ต่อการสนับสนุนการให้บริการ ทำให้ทราบว่าผู้ใช้บริการส่วนใหญ่ มีพฤติกรรมเป็นไปในทิศทางใด สนใจหรือต้องการทรัพยากรสารสนเทศประเภทใด เพื่อจัดหาทรัพยากรสารสนเทศให้ตรงกับความต้องการของผู้ใช้บริการให้มากที่สุด เป็นการสะท้อนให้เห็นถึงความต้องการของผู้ใช้บริการผ่านการใช้คำค้น และสามารถนำไปสู่แนวทางการสร้างสรรบริการ หรือผลิตภัณฑ์ใหม่ ๆ เพื่อการ

บริการในเชิงรุกต่อไป ในการวิจัยครั้งนี้ ผู้วิจัยได้นำเทคนิคการทำเหมืองข้อความ (Text mining) มาใช้ในการวิเคราะห์ความถี่ของคำค้น (Keywords) และแสดงผลในรูปแบบ Visualization เช่น Word clouds และความถี่ เพื่อสามารถทำความเข้าใจกับข้อมูลสรุปได้ง่ายขึ้น (จักรพันธ์ จารุภูมิ, 2554)

วัตถุประสงค์

เพื่อวิเคราะห์คำค้นภาษาอังกฤษที่ผู้ใช้บริการใช้ค้นหาข้อมูลบนระบบโอแพคในรูปแบบ Word cloud

ขั้นตอนและวิธีการดำเนินงาน



1. ก่อนดำเนินการ ขั้นตอนนี้ ประกอบด้วย

1.1 ขั้นตอน Data collection คือ ระบบจะทำการเก็บรวบรวมข้อมูลคำค้น โดยการดึงข้อมูลคำค้นจากฐานข้อมูลโอแพค และจัดเก็บข้อมูลเหล่านั้นไว้ในรูปแบบของไฟล์ข้อความ (Text file)

1.2 ขั้นตอน Data selection เมื่อทำการดึงข้อมูลเรียบร้อยแล้ว ผู้วิจัยจะทำการเลือกข้อมูลที่ต้องการนำมาวิเคราะห์ ในการวิจัยครั้งนี้ เลือกเฉพาะข้อมูลคำค้นที่มีการค้นด้วยเขตข้อมูล ชื่อเรื่อง (คำสำคัญ) ชื่อเรื่อง (คำขึ้นต้น) หัวเรื่อง (คำสำคัญ) และ หัวเรื่อง (คำขึ้นต้น) เนื่องจากเป็นเขตข้อมูลที่มีการใช้บ่อย และฐานข้อมูลโอแพคของห้องสมุด มีการตั้งค่าเริ่มต้นไว้ที่ ชื่อเรื่อง (คำสำคัญ) เป็นอันดับแรก

1.3 ขั้นตอน Data cleaning คือ การทำความสะอาดข้อมูลที่ต้องการนำมาวิเคราะห์ เช่น ลบข้อมูลที่ไม่จำเป็น ข้อมูลที่ไม่ถูกต้อง ข้อมูลภาษาไทย และข้อมูลที่ไม่สมบูรณ์ เป็นต้น

2. กำลังดำเนินการ เป็นขั้นตอนการวิเคราะห์ข้อมูล (Text Analysis) โดยใช้โปรแกรม R ในการวิเคราะห์ข้อมูล มีการเรียกใช้แพ็คเกจ จากโปรแกรม R ที่ชื่อว่าแพ็คเกจ TM (Text Mining Package) เป็นแพ็คเกจหลักในการใช้วิเคราะห์ข้อมูล ซึ่งในขั้นตอนนี้มีกระบวนการย่อย ๆ ดังนี้

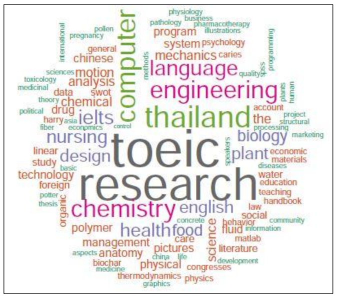
2.1 โหลดข้อมูลคำค้นเข้าไปในคลังข้อมูล (Corpus)

2.2 จัดเตรียมแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ โดยในขั้นตอนนี้มีการเรียกใช้ฟังก์ชัน tm_map จากแพ็คเกจ tm จากนั้นระบบจะทำการลบตัวเลข กำจัดช่องว่าง เครื่องหมายวรรคตอน เปลี่ยนตัวอักษรทั้งหมดจากอักษรพิมพ์ใหญ่เป็นอักษรพิมพ์เล็ก ลบคำหยุด เช่น คำที่ทำหน้าที่เป็นบุพบท คำสันธาน คำสรรพนาม และคำภาษาอังกฤษที่ใช้ทั่วไป (เช่น the, a และ an) เป็นต้น (Jo, Kim, & Shin, 2017)

3. หลังดำเนินการ เป็นขั้นตอนแสดงผลการวิเคราะห์และสรุปข้อมูลคำค้น ในรูปของการแสดงความถี่ของคำในรูปแบบของ Word cloud เพื่อเป็นการสรุปข้อมูลคำค้นที่มีการใช้บ่อยและผู้ใช้มีการค้นถี่ที่สุดในแต่ละช่วงเวลา ซึ่งในขั้นตอนนี้มีการเรียกใช้ฟังก์ชันในแพ็คเกจที่ชื่อว่า Word cloud และ ggplot2 เพื่อการแสดงผลข้อมูล ดังกล่าว

สรุปผล

ภาพที่ 1 ข้อมูลคำค้น เดือน ม.ค. 61



ภาพที่ 2 ข้อมูลคำค้น เดือน ก.พ. 61

ภาพที่ 3 ข้อมูลคำค้น เดือน มี.ค. 61

ตารางที่ 1 ข้อมูลแสดงความถี่ของข้อมูลคำค้นเดือน ม.ค.-มี.ค.61

Words	Frequency	Words	Frequency	Words	Frequency
Toeic	215	Toeic	235	Research	245
Research	194	Research	210	Thailand	214
Nursing	150	Thailand	134	Computer	139
Thailand	146	Computer	122	Nursing	122
Chemistry	124	Chemistry	107	Analysis	121
Cell	115	Engineering	105	Toeic	106
Engineering	114	Language	93	Management	105
Analysis	108	Lelts	85	Language	96
Health	104	Health	83	English	91
Management	104	Nursing	80	Health	90

Words	Frequency	Words	Frequency	Words	Frequency
Mechanics	103	English	75	Chemistry	83
Computer	101	Design	72	Community	83
Biology	93	Plant	70	United	77
Language	84	Food	66	Food	66
Physiology	84	Biology	64	Biology	64

ผลการวิเคราะห์ข้อมูลคำค้นทั้งหมดในรูปของ Word cloud ตั้งแต่เดือนมกราคม 2561 ถึงเดือนมีนาคม 2561 สามารถสรุปได้ว่า คำค้นใดมีขนาดใหญ่แสดงว่ามีความถี่มากในการค้นหา แสดงว่า ผู้ใช้บริการมีความต้องการหรือสนใจในเรื่องนั้น ๆ โดยในเดือนมกราคม 2561 คำค้นที่มีความถี่และผู้ใช้บริการต้องการหรือสนใจมากที่สุด 5 อันดับแรกเรียงตามลำดับความถี่ ได้แก่ Toeic มีความถี่ 215 ครั้ง Research มีความถี่ 194 ครั้ง Nursing มีความถี่ 150 ครั้ง Thailand มีความถี่ 146 ครั้ง และ Chemistry มีความถี่ 124 ครั้ง

เดือนกุมภาพันธ์ 2561 คำค้นที่มีความถี่และผู้ใช้บริการต้องการหรือสนใจมากที่สุด 5 อันดับแรกเรียงตามลำดับความถี่ ได้แก่ Toeic มีความถี่ 235 ครั้ง Research มีความถี่ 210 ครั้ง Thailand มีความถี่ 134 ครั้ง Computer มีความถี่ 122 ครั้ง และ Chemistry มีความถี่ 107 ครั้ง และในเดือนมีนาคม 2561 คำค้นที่มีความถี่และผู้ใช้บริการต้องการหรือสนใจมากที่สุด 5 อันดับแรกเรียงตามลำดับความถี่ ได้แก่ Research มีความถี่ 245 ครั้ง Thailand มีความถี่ 214 ครั้ง Computer มีความถี่ 139 ครั้ง Nursing มีความถี่ 122 ครั้ง และ Analysis มีความถี่ 121 ครั้ง

คำค้นทั้งหมดที่ปรากฏเป็นคำค้นที่มีความถี่และผู้ใช้บริการใช้ในการค้นหามากที่สุด ในช่วงเดือนมกราคมถึงเดือนมีนาคม 2561 จะเห็นได้ว่าทั้งสามเดือนนั้น มีคำค้นบางคำที่ผู้ใช้บริการ ยังมีความสนใจและต้องการอย่างต่อเนื่อง เช่น คำว่า Toeic Thailand และ Chemistry ซึ่งในช่วงดังกล่าวนั้น อาจจะอยู่ในช่วงที่มีการสอบภาษาอังกฤษ Toeic จึงทำให้มีความถี่ของคำค้น Toeic สูงเป็นอันดับแรก และยังคงต้องการในเดือนถัด ๆ ไป และที่สำคัญจากการวิเคราะห์และสรุปข้อมูลคำค้นในครั้งนี้ ทำให้ทราบว่าความถี่ของคำค้นจะเปลี่ยนแปลงไปตามช่วงเวลา แสดงให้เห็นถึงพฤติกรรมความต้องการและความสนใจของผู้ใช้บริการที่เปลี่ยนไปด้วย

อภิปรายผล

การวิเคราะห์และสรุปข้อมูลคำค้นในรูปของ Word cloud และการแสดงข้อมูลความถี่ เพื่อสนับสนุนงานบริการสารสนเทศของห้องสมุดนั้น มีประโยชน์ทั้งต่อผู้ใช้บริการและเจ้าหน้าที่ของห้องสมุด ดังนี้ โดยผู้ใช้บริการสามารถทราบถึงสิ่งที่กำลังได้รับความนิยมหรือค้นหาบ่อยในแต่ละช่วงเวลา จึงควรแสดงผลคำค้นที่มีผู้ใช้บริการค้นหาบ่อยบนเว็บไซต์ เพื่อนำเสนอและนำทางไปยังเรื่องที่สนใจได้ และในส่วนของเจ้าหน้าที่ห้องสมุดนั้น จะทำให้ทราบแนวโน้มความสนใจของผู้ใช้บริการว่ามีความสนใจหรือต้องการทรัพยากรสารสนเทศเรื่องใด หรือสามารถนำข้อมูลเบื้องต้นดังกล่าว เชื่อมโยงกับการจัดหาทรัพยากรสารสนเทศของห้องสมุด เพื่อจัดหาได้ตรงกับความต้องการที่สุด ตัวอย่างเช่น คำค้น Toeic ที่ได้รับความสนใจเยอะในการค้น ทางห้องสมุดสามารถตรวจสอบหรือ เช็คได้ว่าหนังสือเกี่ยวกับ Toeic เล่มไหนมีการยืมหรือใช้เยอะ โดยสามารถทำการจัดหาหนังสือเกี่ยวกับ Toeic ที่มีความทันสมัย เป็นที่นิยมมาให้บริการได้อีกด้วย ซึ่งถือเป็นการอัพเดทข้อมูลทรัพยากรสารสนเทศของห้องสมุดผ่านการวิเคราะห์ข้อมูลคำค้นได้ หรือสามารถนำผลการวิเคราะห์ไปสร้างการวางแผนเพื่อการสร้างความสัมพันธ์กับผู้ใช้บริการ เช่น จัดกิจกรรมที่สอดคล้องกับคำค้นที่ผู้ใช้ให้ความสนใจหรือมีคำค้นที่มีความถี่เยอะ หรือนำข้อมูลสรุปคำค้นที่ได้ไปตกแต่ง

เว็บโอแพคหรือเว็บไซต์ของห้องสมุด เพื่อแสดงให้ผู้ใช้บริการได้ทราบว่าในแต่ละช่วงเวลา ผู้ใช้บริการมีการค้นหรือสนใจเกี่ยวกับเรื่องใด

ข้อเสนอแนะ

1. ควรมีการวิเคราะห์และสรุปข้อมูลคำค้นภาษาไทยในแต่ละช่วงเวลา เพื่อให้ทราบถึงคำค้นภาษาไทยที่ผู้ใช้บริการใช้เยอะจากระบบโอแพค เพื่อเป็นแนวทางในการจัดหาทรัพยากรภาษาไทยที่ตรงกับความต้องการและสนใจของผู้ใช้บริการ
2. การวิเคราะห์และสรุปข้อมูลคำค้นในรูปของ Word cloud นั้น หากมีการแสดงให้ผู้ใช้ทราบผ่านหน้าเว็บโอแพค และสามารถเชื่อมโยงไปยังหนังสือที่มีการใช้เยอะที่เกี่ยวข้องกับคำค้นนั้น ๆ ได้ ก็จะทำให้เว็บโอแพคมีความน่าสนใจมากขึ้นและเป็นประโยชน์กับผู้ใช้บริการด้วย
3. การวิเคราะห์ข้อมูลในครั้งนี้เป็นการวิเคราะห์ข้อมูลคำค้นเป็นคำ ๆ เท่านั้น ยังไม่สามารถวิเคราะห์คำค้นที่เป็นวลี หรือประโยคได้

การนำไปใช้ประโยชน์

1. ทำให้ทราบถึงแนวโน้มความต้องการและความสนใจของผู้ใช้บริการในการค้นหาข้อมูล ทรัพยากรของห้องสมุดว่าในแต่ละช่วงเวลา ผู้ใช้บริการมีความสนใจทรัพยากรประเภทใดมาก โดยห้องสมุดสามารถนำข้อมูลที่ได้จากการวิเคราะห์ครั้งนี้ ใช้ในการวางแผนการจัดกิจกรรมและการจัดหาทรัพยากรสารสนเทศให้สอดคล้องกับความต้องการและความสนใจของผู้ใช้บริการได้
2. การวิเคราะห์ข้อมูลในครั้งนี้ สามารถเป็นเครื่องมือและแนวทางในการช่วยพัฒนาห้องสมุดในเชิงรุก เช่น การจัดแสดงนิทรรศการหนังสือตามคำค้นที่มีการใช้เยอะ การจัดบรรยายเรื่องที่ผู้ใช้บริการสนใจผ่านคำค้น ที่มีการใช้เยอะในแต่ละช่วงเวลา เป็นต้น

รายการอ้างอิง

จักรพันธ์ จารุภูมิ. (2554). *การทำเหมืองข้อความเว็บไซต์สังคมออนไลน์เพื่อวิเคราะห์กระแสการเมืองกรณีศึกษาข้อความสั้นเว็บไซต์ทวิตเตอร์*. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ, คณะเทคโนโลยีสารสนเทศ, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

Castro, R. G. (2015). *RPubs Creating n-gram clouds from text corpora*. Retrieved from http://www.rpubs.com/rgcmme/PLN-09

Jo, Y., Kim, E.-G., & Shin, Y. (2017). *Graphical Keyword Service for Research Papers with Text-Mining Method.* Retrieved from https://doi.org/10.1145/3093241.3093242