

การสร้างแบบจำลองตามหัวข้อจากงานวิจัยด้านบรรณมิติในฐานข้อมูล Scopus
ระหว่างปี ค.ศ. 2019-2023¹

Topic Modeling from Bibliometric Research of the 2019–2023 Data through Scopus

พรณิสา วัฒนศิริ (Pornnisa Wattanasiri)²

วิระพงษ์ จันทร์สนาม (Wirapong Chansanam)³

Received: 06-01-2024

Revised: 19-03-2024

Accepted: 17-04-2024

บทคัดย่อ

บทความนี้เป็นการสร้างแบบจำลองตามหัวข้อของชื่อเรื่องงานวิจัยและดัชนีคำสำคัญของงานวิจัยด้านบรรณมิติจากเอกสารเฉพาะบทความวิจัยที่รวบรวมจากฐานข้อมูล Scopus ในช่วงระหว่างปี ค.ศ. 2019-2023 โดยใช้เทคนิคการสร้างแบบจำลองตามหัวข้อหรือ Topic modeling ด้วยโปรแกรม R studio และ Python โดยเน้นการแสดงผลแผนภาพของคำที่พบมากที่สุดของชื่อเรื่องและดัชนีคำสำคัญ เพื่อให้เห็นคำที่มีความสัมพันธ์หรือคำที่เกิดขึ้นในการวิจัยด้านนี้ ผลการศึกษาพบว่ามีจำนวนเอกสารที่เกี่ยวข้องทั้งหมด 3,931 บทความ และเมื่อนำไปวิเคราะห์ข้อมูลแล้วพบว่า การสร้างแบบจำลองตามหัวข้อของชื่อเรื่องงานวิจัยและดัชนีคำสำคัญ คำที่พบมากที่สุดในการวิเคราะห์ชื่อเรื่องของงานวิจัยด้านบรรณมิตินั้น ส่วนใหญ่เป็นการศึกษาบรรณมิติด้านวิทยาศาสตร์สุขภาพหรือทางการแพทย์ และมีการศึกษาเพิ่มเติมด้านการบริหารจัดการ ด้านสังคมศาสตร์ ด้านการศึกษา และเทคโนโลยีสารสนเทศ เป็นต้น ส่วนผลการวิเคราะห์ดัชนีคำสำคัญที่พบ พบว่าผลการศึกษาไม่แตกต่างกันจากผลการศึกษาในด้านชื่อเรื่องของงานวิจัย และมีการศึกษาเพิ่มเติมด้านสิ่งแวดล้อมด้วย ซึ่งการศึกษานี้เป็นประโยชน์สำหรับนักวิชาการ นักวิจัย หรือผู้ที่สนใจการวิเคราะห์งานวิจัยด้วยการวิเคราะห์แบบจำลองหัวข้อประเด็นอื่นหรือเทคนิคอื่นๆ ได้

คำสำคัญ: การสร้างแบบจำลองตามหัวข้อ บรรณมิติ การวิเคราะห์ข้อมูล การสร้างแผนภาพความสัมพันธ์

¹ บทความวิจัยนี้เป็นส่วนหนึ่งที่ขยายผลการศึกษาเพิ่มเติมของปริญญาโทสาขาสารสนเทศศาสตรมหาบัณฑิต สาขาวิชาสารสนเทศศาสตร์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น: This article is an extended version of Thesis in Master Degree of Information Science, Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University

² นักศึกษาระดับปริญญาโท สาขาวิชาสารสนเทศศาสตร์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น: Master's degree student, Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, E-mail: pornnisa.w@kkumail.com

³ รองศาสตราจารย์ และหัวหน้าสาขาวิชาสารสนเทศศาสตร์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น: Assoc. Prof., & Head of Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, E-mail: wirach@kku.ac.th

Abstract

This article aims to analyze the titles and indexed keywords of research articles in bibliometric research using topic modeling techniques. The analysis was based on articles collected from the Scopus database between 2019 and 2023, focusing on identifying the most frequently occurring words in the titles and indexed keywords of the articles to reveal the relationships or occurring words among the topics studied in this field in R Studio and Python. The study identified 3,931 relevant articles and topic models based on the most frequently occurring words within the analyzed titles and indexed keywords. These topics primarily encompassed health sciences and medicine, with additional management, social sciences, education, and information technology coverage. The study also revealed that the most commonly indexed keywords did not differ significantly from the frequently occurring words in the analyzed titles, further emphasizing environmental studies. This study was valuable for researchers, academics, and others interested in analyzing research articles using topic modeling or other related techniques.

Keyword: Topic modeling, Bibliometric, Data analysis, Mapping Science

บทนำ

บรรณมิติ (Bibliometrics) คือ กระบวนการที่ใช้การวิเคราะห์ทางสถิติในการสกัดข้อมูลบรรณานุกรมที่สามารถวัดได้จากการผลงานตีพิมพ์ และวิเคราะห์ว่าองค์ความรู้ภายในการตีพิมพ์นั้นถูกนำไปใช้ในทิศทางใดบ้าง (Agarwal et al., 2016) เช่น จำนวนการอ้างอิงผลงานวิจัย ปัจจัยที่มีผลต่อผลกระทบของงานวิจัย ค่า h-index ของผู้เขียน ประเทศ หรือสถาบันที่ผลิตผลงานวิจัย เป็นต้น การวิเคราะห์แบบชี้วัดของผลกระทบเหล่านี้สามารถช่วยให้นักวิจัยระบุรูปแบบและแนวโน้มในวรรณกรรมสิ่งพิมพ์ การประเมิน การวัดผลผลิตของนักวิจัย และระบุพื้นที่วิจัยที่ต้องการการศึกษาเพิ่มเติม ซึ่งมีงานวิจัยของ Zupic & Čater (2015) ใช้เทคนิคการวิเคราะห์บรรณมิติและการวัดผลของงานวิจัยหรืองานที่เกี่ยวข้องกับการจัดการและองค์กรโดยใช้ข้อมูลจากเอกสารทางวิชาการ เช่น บทความวิจัย บทความวิชาการ หรือ สิ่งพิมพ์ทางวิชาการ นำมาวิเคราะห์ผลต่างๆ เช่น การวิเคราะห์การอ้างอิง (citation analysis), การวิเคราะห์การอ้างอิงร่วม (co-citation analysis), การวิเคราะห์การจับคู่บรรณานุกรม (bibliographic coupling) และ การวิเคราะห์เหมืองข้อความ (text mining) เพื่อสร้างความเข้าใจและการวิเคราะห์ข้อมูลที่มีประโยชน์ต่อการพัฒนาการวิจัยและการจัดการองค์กรในอนาคต

การวิเคราะห์บรรณมิติได้รับความนิยมสูงขึ้นในการนำมาเป็นเทคนิคหนึ่งในการวิจัยอย่างหลากหลายสาขาวิชาทั่วโลก (Ellegaard & Wallin, 2015) ซึ่งเกี่ยวข้องกับการวิเคราะห์การตีพิมพ์ผลงานวิจัยเพื่อให้ได้ข้อมูลเชิงลึกเกี่ยวกับแนวโน้มของการวิจัยที่มีปริมาณมาก (Donthu et al., 2021) รวมถึงแนวโน้มของการทำงานร่วมกัน และการประเมินผลกระทบ การวิเคราะห์บรรณมิติช่วยให้นักวิจัยได้รับความเข้าใจอย่างมีประสิทธิภาพ และอย่างเป็นรูปธรรมเกี่ยวกับการพัฒนาและลักษณะของสาขาวิชาวิจัยที่ต้องการศึกษาได้ อีกทั้ง การวิเคราะห์บรรณมิติทำให้ทราบข้อมูลเชิงสถิติและแผนภาพเพื่อสรุปแนวโน้มในบริบทของการวิจัยที่เลือกไว้ (Pritchard, 1969) ซึ่งสามารถเปิดเผยแนวโน้มการวิจัยในสาขาที่กำหนดได้ เช่น ผลลัพธ์จากการเผยแพร่งานวิจัย ประสิทธิภาพของผู้เขียน การกระจายตำแหน่งทางภูมิศาสตร์ของการเผยแพร่ ความร่วมมือทางวิทยาศาสตร์ และการเปลี่ยนแปลงของประเด็นร้อนแรงในช่วงเวลาต่างๆ (Liu et al., 2011)

ขณะเดียวกันงานวิจัยด้านบรรณมิติที่มีการเผยแพร่ผลงานตีพิมพ์อย่างมากในฐานข้อมูลสากลขนาดใหญ่ อย่างเช่น Scopus, Web of Science และ Dimensions เป็นต้น การสร้างแบบจำลองตามหัวข้อที่มีการวิจัยทางด้านบรรณมิติจะช่วย

ทำให้เกิดประโยชน์แก่ชุมชนนักวิจัยต่าง ๆ ที่ศึกษาได้จากชื่อเรื่องและดัชนีคำสำคัญของผู้เขียนหรือนักวิจัยที่มีต่อผลงานด้านบรรณมิติได้ ดังตัวอย่างงานวิจัยของ Bedekar and Desai (2022) ศึกษาวิเคราะห์โดยใช้ชื่อเรื่องของงานวิจัยที่มีคำสำคัญจากคำค้นว่า ‘Covid-19’ หรืองานวิจัยโดย Garcia-Zorita and Pacios (2018) วิเคราะห์แบบจำลองหัวข้อจากชื่อเรื่องของเอกสารด้านศิลปะ Mudéjar และงานวิจัยของ Natukunda & Muchene (2023) ใช้เทคนิค Latent Dirichlet Allocation (LDA) เพื่อสร้างหัวข้อที่เป็นตัวแทนจากชื่อเรื่องและบทคัดย่อของเอกสารสำหรับการทบทวนวรรณกรรมเชิงระบบ (systematic review) เป็นต้น

จากความสำคัญของการวิเคราะห์ข้อมูลด้วยเทคนิคหรือแนวทางบรรณมิติดังที่กล่าวมาแล้วนั้น ผู้วิจัยจึงสนใจที่จะศึกษาด้วยเทคนิคหรือเครื่องมือที่สำคัญที่ช่วยในการวิเคราะห์ทางคณิตศาสตร์เกี่ยวกับฐานข้อมูลวรรณกรรมที่นักวิจัยต้องการได้ด้วยการใช้ภาษา Python ในการวิเคราะห์ข้อมูล (Saxton, 2018) ใช้โปรแกรม R studio โดยใช้ภาษา R ในการวิเคราะห์ข้อมูล (Grün & Hornik, 2011; Grün & Hornik, 2023) ได้แก่ การสร้างแบบจำลองตามหัวข้อ หรือ Topic modeling ซึ่งเป็นเทคนิคการทำ text mining เพื่อสกัดโครงสร้างการศึกษาความหมาย (semantics) ที่ซ่อนอยู่ภายในชุดข้อมูลเอกสารหลายๆ เอกสารได้อย่างอัตโนมัติ (Guo et al., 2017)

การศึกษาวิจัยที่ผ่านมา มีการสร้างแบบจำลองตามหัวข้อในการทำวิจัย ซึ่งมีส่วนช่วยให้สามารถแยกแยะหัวข้อหรือเรื่องที่อยู่ในเอกสารหรือข้อมูลได้โดยไม่ต้องอ่านทั้งหมดและประหยัดเวลาในการวิเคราะห์ข้อมูล นอกจากนี้ แบบจำลองหัวข้อยังช่วยในการสกัดความสัมพันธ์ระหว่างคำ หรือความสัมพันธ์ระหว่างหัวข้อที่ต่างกันได้อย่างไรก็ตาม ในมุมมองธุรกิจหรือผู้ใช้งานสามารถนำปรับปรุงและพัฒนาระบบงานต่างๆ ได้ดียิ่งขึ้น โดยงานวิจัยตัวอย่างที่พบ เช่น Yeo Jin & Youngmin (2023) สามารถนำแนวคิดแบบจำลองหัวข้อมาใช้เพื่อช่วยในการวิเคราะห์การตลาดและจัดการสินค้า ในขณะที่ Choi & Um (2023) นำเสนอวิธีการใช้แบบจำลองหัวข้อในการวิเคราะห์ข่าวเกี่ยวกับ Covid-19 รวมถึงงานวิจัยการวิเคราะห์รีวิวร้านอาหารบน Yelp หรือ TripAdvisor โดยใช้เทคนิคการสร้างแบบจำลองตามหัวข้อ และการวิเคราะห์อารมณ์ความรู้สึก (sentiment analysis) มาช่วยในการวิเคราะห์ความคิดเห็นของผู้ใช้งานเพื่อปรับปรุงและพัฒนาระบบการบริการให้ดียิ่งขึ้นได้ (Zhang et al., 2022) จึงเห็นได้ว่า การศึกษาที่ผ่านมากการวิเคราะห์แบบจำลองหัวข้อสามารถนำมาประยุกต์ใช้ในทางการศึกษาได้หลากหลายรูปแบบทั้งการวิเคราะห์หัวข้อทางการวิจัยในหลากหลายประเด็น รวมถึงการค้นหาคำปัญหาหรือ ข้อคิดเห็นที่เกี่ยวข้อง ซึ่งทำให้ผู้ที่มีส่วนเกี่ยวข้องสามารถนำข้อมูลที่ได้ไปใช้ในการแก้ไขปัญหา พัฒนาการจัดการเรียนการสอน หรือพัฒนาหัวข้อในการดำเนิน การวิจัยทางการศึกษาเพื่อเติมเต็มช่องว่างหรือต่อยอดความรู้ ในอนาคตได้

อย่างไรก็ตาม งานวิจัยในครั้งนี้คือ การสร้างแบบจำลองตามหัวข้อจากงานวิจัยด้านบรรณมิติจากฐานข้อมูล Scopus ระหว่างปีค.ศ. 2019-2023 โดยฐานข้อมูล Scopus นั้นเป็นฐานข้อมูลบรรณานุกรมขนาดใหญ่ที่ครอบคลุมงานวิชาการที่มีคุณภาพหลากหลายสาขาวิชา (Chander et al., 2022; Donthu et al., 2021; Sweileh, 2020) ส่วนโปรแกรม R-studio และ Python ที่เลือกใช้ในการวิเคราะห์ครั้งนี้เพื่อให้การสร้างแบบจำลองตามหัวข้อมีความชัดเจนยิ่งขึ้น เนื่องจากภาษา R เป็นภาษาโปรแกรมที่ได้รับความนิยมในการทำงานด้านการวิเคราะห์ข้อมูลและสถิติ มีชุดคำสั่งและไลบรารี (library) ที่ช่วยให้ผู้ใช้สามารถดำเนินการวิเคราะห์ข้อมูลได้อย่างหลากหลายในการจัดการข้อมูล และ Python เป็นภาษาโปรแกรมที่มีความยืดหยุ่นและมีประสิทธิภาพสูงในการประมวลผลข้อมูล โดยทั้งสองโปรแกรมนี้มีความสามารถในการสร้างกราฟ และ แผนภาพความสัมพันธ์ (visualization) ได้อย่างมีประสิทธิภาพ รองรับได้หลายระบบปฏิบัติการ และเป็นโปรแกรมฟรีไม่เสียค่าใช้จ่าย (Deshmukh et al., 2019) ซึ่งจากการศึกษาที่ผ่านมาในการวิเคราะห์แบบจำลองหัวข้อในประเด็นอื่น ๆ ยังไม่พบการศึกษาแบบจำลองหัวข้อด้านบรรณมิติ ซึ่งจะเป็นประโยชน์สำหรับนักวิชาการ นักวิจัย หรือผู้ที่สนใจการวิเคราะห์งานวิจัยด้วยการวิเคราะห์แบบจำลองหัวข้อด้วยเทคนิควิธีการต่าง ๆ มากยิ่งขึ้น อีกทั้งเป็นเครื่องมือที่สามารถนำมาใช้ในการศึกษาและวิจัยได้

ในหลายมิติ และทำให้ทราบถึงทิศทางหรือรูปแบบแนวโน้มของเอกสารงานวิจัยด้านบรรณมิติที่เกิดจากการสร้างแบบจำลองด้วยหัวข้อจากชื่อเรื่องและดัชนีคำสำคัญของงานวิจัยนี้

วัตถุประสงค์ของการวิจัย

เพื่อวิเคราะห์และสร้างแผนภาพความสัมพันธ์ จากชื่อเรื่องและดัชนีคำสำคัญของงานวิจัยด้านบรรณมิติในช่วงระหว่างปี ค.ศ. 2019-2023 จากฐานข้อมูล Scopus โดยแสดงผลแบบจำลองหัวข้อด้วยเทคนิคการสร้างแบบจำลองตามหัวข้อในโปรแกรม R Studio และ Python

กรอบแนวคิดในการวิจัย



ภาพประกอบ 1 กรอบแนวคิดการวิจัย (ประยุกต์จากแหล่งที่มา: Isoaho et al., 2021)

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิเคราะห์งานวิจัยเพื่อการสร้างแบบจำลองตามหัวข้อ (Topic modeling) โดยมีขั้นตอนการวิจัยประกอบด้วย 4 ขั้นตอนหลัก ดังนี้

1. ขั้นตอนที่ 1 การกำหนดหัวข้อและการรวบรวมข้อมูล

ในการวิจัยครั้งนี้ กำหนดหัวข้อเพื่อการรวบรวมข้อมูล คือ งานวิจัยด้านบรรณมิติในฐานข้อมูล Scopus โดยผู้วิจัยใช้เทคนิคในการสืบค้นเอกสารที่เกี่ยวข้องในฐานข้อมูล Scopus เมื่อวันที่ 4 มีนาคม 2566 ดังนี้ (TITLE-ABS-KEY ("bibliometrix" or "scientometrics" or "bibliometric") AND PUBYEAR > 2019 AND PUBYEAR < 2022 AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (LANGUAGE,"English")) AND (LIMIT-TO (EXACTKEYWORD, "Bibliometric") OR LIMIT-TO (EXACTKEYWORD, "Bibliometrics"))) เลือกประเภทเอกสารเฉพาะ บทความวิจัย และภาษาอังกฤษเท่านั้น กำหนดเลือกช่วงระยะเวลาปีค.ศ. 2019 - 2023 จากนั้นดึงข้อมูลเพื่อเก็บบันทึกในรูปแบบไฟล์นามสกุล .csv จึงมีเอกสารที่เกี่ยวข้องทั้งหมด 3,931 บทความ

2. ขั้นตอนที่ 2 การทำความสะอาดข้อมูล

จากขั้นตอนที่ 1 ผู้วิจัยนำไฟล์ .csv ที่ได้บันทึกไว้มาทำความสะอาดข้อมูลด้วยระบบมือ (Manual system) โดยตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล แก้ไขหรือลบข้อมูลที่ไม่ถูกต้องหรือซ้ำกัน หลังจากนั้นนำไฟล์ดังกล่าวเข้าวิเคราะห์ข้อมูลด้วยโปรแกรม R studio เวอร์ชัน 4.2.2 เพื่อสร้างแบบจำลองหัวข้อ (Topic modeling) ด้วย package text mining (tm) จากนั้นลบคำที่ไม่มีประโยชน์ต่อการวิเคราะห์หรือคำฟุ่มเฟือย (Stop-word) หรือคำที่ไม่เกี่ยวข้องออก เช่น the, a, an, in, of เป็นต้น ต่อมาเลือกวิเคราะห์ข้อมูลในคอลัมน์เฉพาะ ชื่อเรื่อง (Titles) และดัชนีคำสำคัญ (Indexed Keywords) จากนั้นเลือกกำหนดค่าผลลัพธ์จำนวนการแสดงผลคำ (Term) $n = 20$ ที่พบมากที่สุด 20 อันดับแรก โดยการจัดการแสดงผลคำที่พบมากที่สุด 20 อันดับแรก มีผลในการช่วยสนับสนุนการตัดสินใจทางวิชาการ ให้ผู้วิจัยหรือนักวิจัยเข้าใจแนวโน้ม

และทิศทางการวิจัยของเรื่องที่กำลังศึกษาได้มากขึ้น และโทเค็นคำ (Word tokens) กำหนดจำนวน $n > 20$ เพื่อเตรียมการ
สร้างแบบจำลองตามหัวข้อเป็นแผนภาพข้อมูล (Data visualization) ของงานวิจัยด้านบรรณมิติ และวิเคราะห์การวัดผลการ
แบ่งกลุ่ม LDA ด้วยเทคนิค Coherence metrics ในโปรแกรม Python ซึ่งเป็นวิธีการวัดความสมเหตุสมผลของการสร้าง
แบบจำลอง โดยใช้ความเชื่อมโยงระหว่างคำในเซตของคำที่เกี่ยวข้องกันในเอกสาร เพื่อประเมินความชัดเจนและความ
สมเหตุสมผลของกลุ่มคำที่สร้างขึ้นโดยอัตโนมัติ

3. ขั้นตอนที่ 3 สร้างแผนภาพข้อมูล

ผู้วิจัยสร้างรายงานผลการวิเคราะห์ด้วยการสร้างแผนภาพ (Visualization) โดยใช้โปรแกรม R studio และ Python
เนื่องจากทั้งสองโปรแกรมนี้นี้มีประสิทธิภาพสูงสำหรับการทำงานทางด้านสถิติสมัยใหม่ โดยเฉพาะงานด้านการวิเคราะห์ข้อมูล
(data analysis) และด้านวิทยาการข้อมูล (data science) ในขั้นตอนนี้เป็นการสร้างแบบจำลองตามหัวข้อของชื่อเรื่อง
งานวิจัยและดัชนีคำสำคัญงานวิจัยด้านบรรณมิติแสดงผลด้วยแผนภาพ และใช้โปรแกรม Python ในการวิเคราะห์แบบจำลอง
หัวขื่อนำเสนอแผนภาพด้วย pyLDAVis package (Sievert & Shirley, 2014)

4. ขั้นตอนที่ 4 วิเคราะห์การแปลผลและสรุป

จากขั้นตอนที่ 3 รายงานผลการวิเคราะห์ที่ได้จากโปรแกรมเพื่อการวิเคราะห์ การแปลผล และสรุปผลการวิจัย โดย
แบ่งการวิเคราะห์ผล 5 หัวข้อ ได้แก่ 1) ผลการวิเคราะห์แผนภาพแบบจำลองหัวข้อ: ชื่อเรื่อง 2) ผลการวิเคราะห์แผนภาพ
แบบจำลองหัวข้อ: ดัชนีคำสำคัญ 3) ผลการวิเคราะห์แผนภาพความสัมพันธ์ของโทเค็นคำ (word tokenization): ชื่อเรื่อง 4)
ผลการวิเคราะห์แผนภาพความสัมพันธ์ของโทเค็นคำ (word tokenization): ดัชนีคำสำคัญ และ 5) ผลการวิเคราะห์การ
แบ่งกลุ่ม Latent Dirichlet Allocation (LDA) และแผนภาพด้วย pyLDAVis

สรุปผลการวิจัย

1. ผลการวิเคราะห์แผนภาพแบบจำลองหัวข้อ: ชื่อเรื่อง

จากการนำไฟล์ข้อมูลโดยมี entity ที่สำคัญ ได้แก่ Titles และ Indexed Keywords ในภาพประกอบ 2 เข้า
วิเคราะห์ในโปรแกรม R เพื่อแสดงผลการวิเคราะห์ตาม ภาพประกอบ 3 ผลการศึกษาพบว่า การสร้างแบบจำลองหัวข้อด้วย
เทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling)

กลุ่มที่ 1 พบคำที่มากที่สุด ดังนี้ trends: แนวโน้ม, 100: ตัวเลข 100, knowledge: ความรู้, literature:
วรรณกรรม, based: ที่มาหรือพื้นฐาน, covid-19: โรคโควิด-19, between: ระหว่าง, web: เว็บไซต์, management: การ
จัดการ, academic: วิชาการ, review: การตรวจสอบ, visualized: แสดงผลในรูปแบบกราฟิกหรือภาพถ่าย, information:
ข้อมูล, bibliometrics: บรรณมิติ, impact: ผลกระทบ, international: ระดับนานาชาติ, mapping: แผนที่, medicine:
การแพทย์, output: ผลลัพธ์, systematic: อย่างเป็นระบบ

กลุ่มที่ 2 พบคำในชื่อเรื่องที่มีมากที่สุด ดังนี้ global: ระดับโลก, scientific: ทางวิทยาศาสตร์, cited: การถูกอ้างอิง,
trends: แนวโน้ม, review: การตรวจสอบ, literature: วรรณกรรม, health: สุขภาพ, during: ระหว่าง, development:
การพัฒนา, years: ปี, covid-19: โรคโควิด-19, field: พื้นที่, mapping: แผนที่, hotspots: จุดเด่น, based: พื้นฐาน,
cancer: โรคมะเร็ง, education: การศึกษา, management: การจัดการ, top: อันดับสูงสุด, social: สังคม

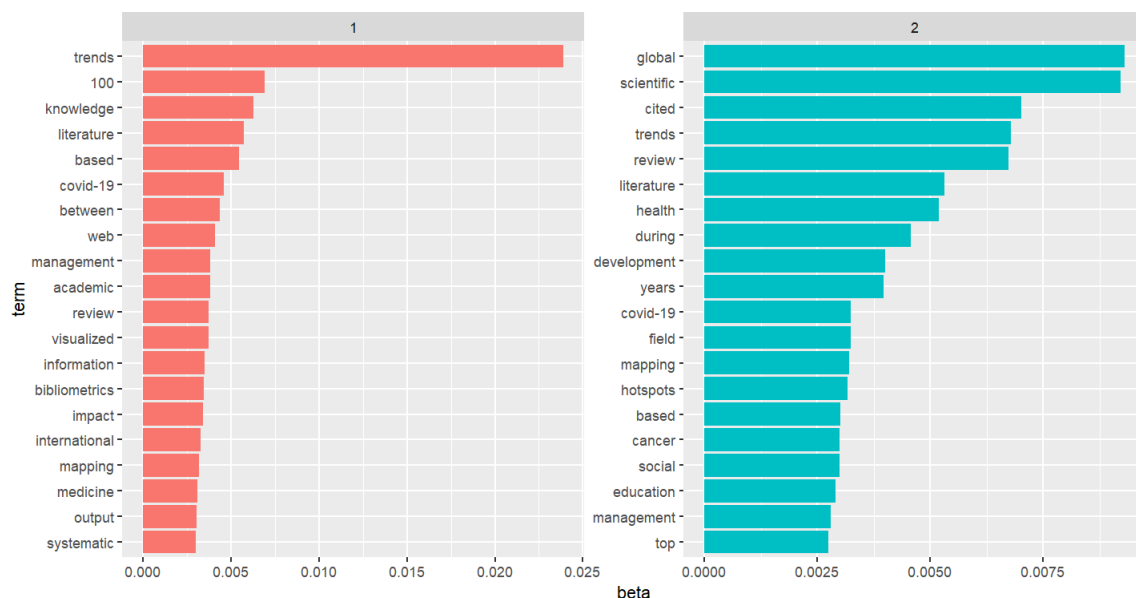
โดยสรุปได้ว่า ชื่อเรื่องบทความวิจัยในงานวิจัยด้านบรรณมิติ คำที่พบมากที่สุดคือ คำว่า trends: แนวโน้ม (กลุ่มที่ 1)
และคำว่า global: ทัวโลก (กลุ่มที่ 2) ทำให้ชี้ชัดว่าชื่อเรื่องของงานวิจัยด้านบรรณมิตินิยมศึกษาเกี่ยวกับแนวโน้มการวิจัย และ
ผลงานวิจัยบรรณมิติในทั่วโลกกระดบสากล เป็นต้น ในขณะเดียวกัน โดยแบบจำลองตามหัวข้อส่วนใหญ่งานวิจัยด้านบรรณมิติ
เกี่ยวข้องกับทางการวิทยาศาสตร์สุขภาพหรือทางการแพทย์ และด้านการบริหารจัดการหรือสังคมศาสตร์ ดังคำที่ปรากฏใน

ภาพประกอบ 3 ซึ่งมีค่าที่พบแล้วมีความโดดเด่นกว่าคำอื่นๆ เช่น trends: แนวโน้ม, knowledge: ความรู้, global: ทั่วโลก, cited: การอ้างอิง, covid-19: โรคโควิด-19 เป็นต้น อีกทั้ง เนื่องจากการระบาดของโรค Covid-19 ปรากฏตัวครั้งแรกในเดือนธันวาคมของปี ค.ศ.2019 ในประเทศจีนทำให้เกิดการแพร่กระจายของเชื้อไปยังผู้คนทั่วโลก (Wu et al., 2020) ส่งผลให้การวิจัยด้านบรรณมิติมีค่าที่เกี่ยวข้องกับโรคระบาด Covid-19 ด้วยเช่นกัน ดังคำที่พบ เช่น health: สุขภาพ, covid-19: โรคโควิด-19, cancer: โรคมะเร็ง เป็นต้น

ภาพประกอบ 3 แสดงถึง การสร้างแบบจำลองหัวข้อด้วยเทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) ในการหาคำที่ปรากฏในชื่อเรื่องที่เกี่ยวข้องกับบทความวิจัย โดยชื่อเรื่องบทความวิจัยส่วนใหญ่เกี่ยวข้องกับทางวิทยาศาสตร์สุขภาพหรือทางการแพทย์ และด้านการบริหารจัดการหรือสังคมศาสตร์ แต่ด้วยสถานการณ์โรคระบาด covid-19 ทำให้การวิจัยด้านบรรณมิติมีค่าที่เกี่ยวข้องกับโรคระบาด covid-19 มากขึ้น เช่น โรคโควิด-19 และระหว่างสถานการณ์โรคระบาดนี้ยังพบว่า มีการวิจัยด้านบรรณมิติที่เกี่ยวข้องกับการแพร่ระบาดของโรคนี้เพิ่มขึ้นอย่างมาก อย่างไรก็ตาม การวิจัยด้านบรรณมิติยังคงมีความสำคัญและน่าสนใจในด้านอื่นๆ ด้วย เช่น การบริหารจัดการ และการศึกษา ที่แสดงให้เห็นถึงความหลากหลายในการประยุกต์ใช้เทคนิค Topic modeling ในการวิเคราะห์บทความวิจัยในหลายสาขาวิชา

	A	B	C	D	E	F	G	H	I	J	K	L
1	Authors	Author full	Author(s)	Titles	Year	Indexed Keywords						
2	Chuang H.	Chuang, H	57848823	The use of the time-to-event index (Tevent) to compare the negative impact of COVID-19 on public health among continents/regions in 2020	2022	COVID-19; Disease Outbreaks; Humans; Pandemics; Public Health; SARS-CoV-2; algorithm						
3	Xue J.; Zh	Xue, Jian	55262040	A review of supply chain coordination management based on bibliometric data	2022	Data visualization; Supply chains; Sustainable development; Time series analysis; Bibliometrics						
4	Chang C.;	Chang, Cu	58019795	Bibliometric analysis of the global research trends and hotspots in chondroma from 2000 to 2020	2022	Aggression; Bibliometrics; Chondroma; Choriocarcinoma; Female; Humans; Research Perspectives						
5	Guzmán-V	Guzmán-V	55808771	Celebrating Higher Education's 50 years: trends, absences, and opportunities	2022							
6	Zhang R.;	Zhang, Ra	57430499	Scientometric analysis of kidney disease and gut microbiota from 2001 to 2020 based on Web of Science	2022	Databases, Factual; France; Gastrointestinal Microbiome; Humans; Indian; Renal Insufficiency						
7	Xi M.; Gac	Xi, Mengd	57979875	Bibliometric Analysis of Research Relating to IgA Nephropathy from 2010 to 2021	2022	Authorship; Bibliometrics; Databases, Factual; Humans; Immunoglobulin A; Universities; Immunology						
8	D'Aniello	D'Aniello	57243936	Academic Health Centers' configurations, scientific productivity, and impact: Insights from the Italian setting	2022	Academic Medical Centers; Humans; Organizations; article; bibliometrics; cluster analysis; Dispersions						
9	Chen D.;	Chen, Der	57877467	Dispersion and Disparity: Bibliometric and Visualized Analysis of Research on Climate Change Science Communication	2022	Bibliometrics; Climate Change; Interdisciplinary Communication; climate change; decentralization						
10	Naeini A.E	Naeini, Ali	57222347	Conceptual structure and perspectives on "innovation management": A bibliometric review	2022	Maps; Patents and inventions; Bibliometric; Bibliometric techniques; Bibliometrics analysis						
11	Lin F.; Ch	Lin, Feihoi	58038652	A bibliometric analysis of autophagy in lung diseases from 2012 to 2021	2022	Autophagy; Bibliometrics; COVID-19; Humans; Hypertension; Pulmonary; Lung Neoplasms						

ภาพประกอบ 2 ตัวอย่างไฟล์ .csv ที่นำไปวิเคราะห์ข้อมูล



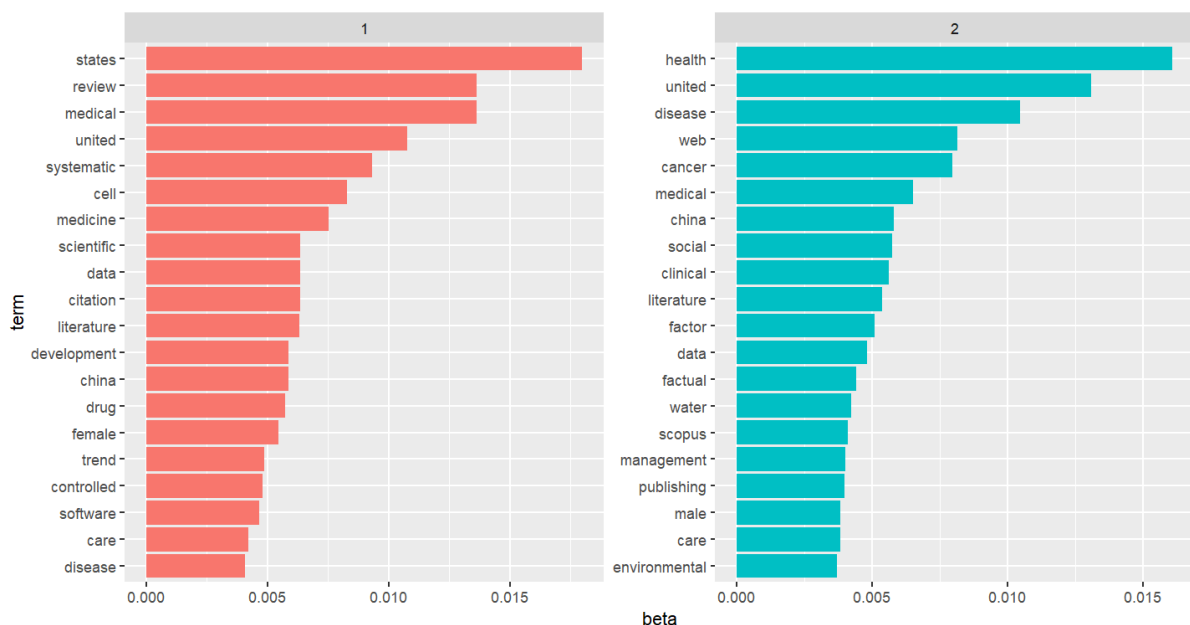
ภาพประกอบ 3 คำที่พบมากที่สุด 20 คำแรกของงานวิจัยด้านบรรณมิติ (ชื่อเรื่อง)

2. ผลการวิเคราะห์แผนภาพแบบจำลองหัวข้อ: ดัชนีคำสำคัญ

ภาพประกอบ 4 แสดงผลการวิเคราะห์ดัชนีคำสำคัญ ได้ดังนี้ กลุ่มที่ 1 states: รัฐ, review: การตรวจสอบ, medical: การแพทย์, united: สหรัฐอเมริกา, systematic: อย่างเป็นระบบ, cell: เซลล์, medicine: ทางการแพทย์, scientific: ทางวิทยาศาสตร์, citation: การอ้างอิง, literature: วรรณกรรม, development: การพัฒนา, china: จีน, drug: ยา, female: เพศหญิง, trend: แนวโน้ม, controlled: การควบคุม, software: ซอฟต์แวร์, care: การดูแล, disease: โรค, data: ข้อมูล

กลุ่มที่ 2 ได้แก่ คำว่า health: สุขภาพ, united: สหรัฐอเมริกา, disease: โรค, web: เว็บไซต์, cancer: โรคมะเร็ง, medical: การแพทย์, china: จีน, social: ด้านสังคม, clinical: ทางคลินิก, literature: วรรณกรรม, factor: ปัจจัย, data: ข้อมูล, factual: ตามความเป็นจริง, water: น้ำ, scopus: ฐานข้อมูล scopus, management: การจัดการ, publishing: การเผยแพร่, male: เพศชาย, care: การดูแล, environmental: สิ่งแวดล้อม

โดยสรุปได้ว่า ดัชนีคำสำคัญที่พบมากที่สุดคือ คำว่า states: รัฐ (กลุ่มที่ 1) และคำว่า health: สุขภาพ (กลุ่มที่ 2) อาจพิจารณาได้ว่า ดัชนีคำสำคัญของงานวิจัยด้านบรรณมิตินั้น มีการศึกษาบรรณมิติเกี่ยวกับการระบุพื้นที่ในเขตหรือรัฐต่างๆ ที่กำหนดในการวิจัย และการวิจัยบรรณมิติเกี่ยวกับด้านสุขภาพมากที่สุด เป็นต้น โดยการศึกษาวิจัยด้านบรรณมิติส่วนใหญ่เกี่ยวข้องกับด้านวิทยาศาสตร์สุขภาพและการแพทย์ โดยมีดัชนีคำสำคัญที่สื่อความหมายเกี่ยวข้องกับงานวิจัยทางการแพทย์ เช่น health: สุขภาพ, medical : การวิจัยทางการแพทย์, cancer: โรคมะเร็ง, drug: ยา, clinical: การทดลองทางคลินิก, care: การดูแล และ disease: โรค เป็นต้น นอกจากนี้ ยังมีคำที่เกี่ยวข้องกับการบริหารจัดการและเทคโนโลยีสารสนเทศ เช่น development: การพัฒนา, management: การจัดการ, controlled: การควบคุม, software: ซอฟต์แวร์, data: ข้อมูล, web: เว็บไซต์ และด้านสิ่งแวดล้อม เช่น environmental: สิ่งแวดล้อม, water: น้ำ รวมถึงด้านสังคมศาสตร์ เช่น social: สังคม, publishing: การเผยแพร่ตีพิมพ์, literature: วรรณกรรม, systematic: อย่างเป็นระบบ, citation: การอ้างอิง เป็นต้น ดังนั้น การศึกษาวิจัยด้านบรรณมิติเกี่ยวข้องกับด้านวิทยาศาสตร์สุขภาพและการแพทย์มากที่สุด โดยมีคำที่เกี่ยวข้องกับงานวิจัยทางการแพทย์และสุขภาพอย่างมาก และในขณะเดียวกันยังเกี่ยวข้องกับด้านการบริหารจัดการและเทคโนโลยีสารสนเทศ และด้านสิ่งแวดล้อม (ภาพประกอบ 4)



ภาพประกอบ 4 คำที่พบมากที่สุด 20 คำแรกของงานวิจัยด้านบรรณมิติ (ดัชนีคำสำคัญ)

3. ผลการวิเคราะห์แผนภาพความสัมพันธ์ของโทเคนคำ (word tokenization): ชื่อเรื่อง

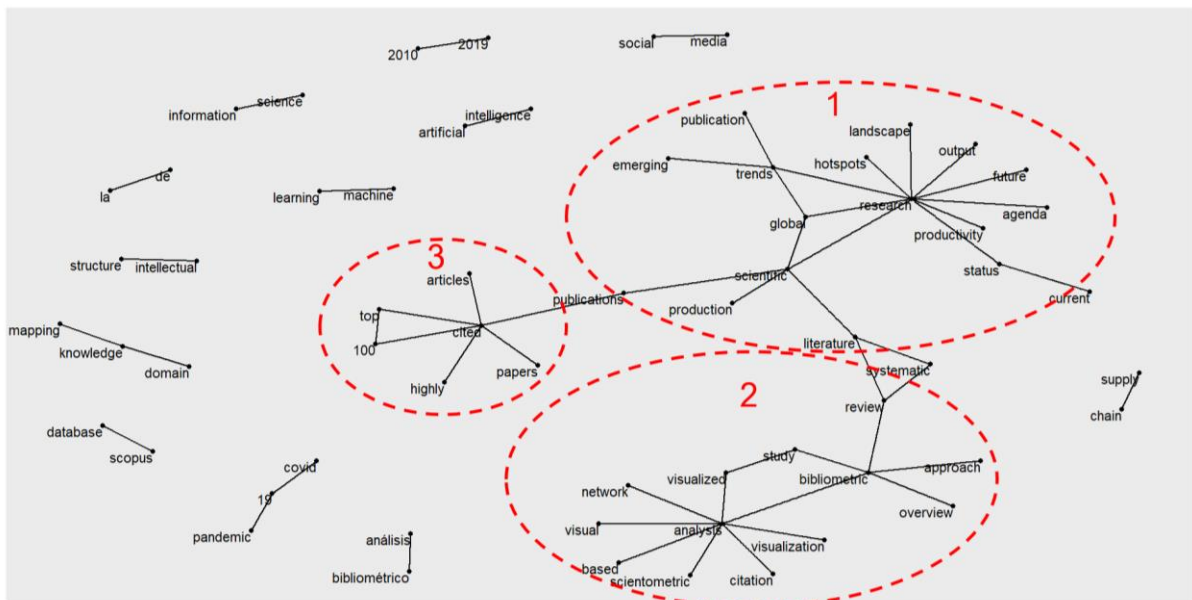
ภาพประกอบ 5 การวิเคราะห์หาความสัมพันธ์ของชื่อเรื่องโดยการทำโทเคนคำ ซึ่งเป็นกระบวนการแยกชิ้นส่วนของข้อความออกเป็นหน่วยเล็ก ๆ เรียกว่าโทเคน (tokens) ใช้เพื่อบันทึกจำนวนคำแต่ละคำ และวิเคราะห์คุณสมบัติของคำ เช่น ความถี่ของคำที่ปรากฏในเอกสารหรือข้อความ, จำนวนคำทั้งหมดในเอกสารหรือข้อความ, ความยาวของคำ และความหมายของคำ โดยการศึกษาครั้งนี้ใช้ visualization คำสั่ง library(ggraph) ในภาษา R เพื่อแสดงผลเป็นกราฟเครือข่าย (network graph) ของ bigram (คู่คำ) ที่ได้จากการนับความถี่ของคำในเอกสาร เพื่อให้เห็นแบบจำลองตามหัวข้อและลักษณะของเครือข่ายได้ โดยจัดวางโหนดหรือจุดที่มีความสัมพันธ์ของคู่คำให้มีความสมดุลกัน และแสดงเส้นเชื่อมแทนความสัมพันธ์ระหว่างคู่คำ เพื่อให้เห็นความสัมพันธ์ระหว่างคำได้อย่างชัดเจน ดังภาพประกอบ 5 ผู้วิจัยกำหนดกลุ่มวงกลมสีแดงออกเป็นทั้งหมด 3 กลุ่มใหญ่ เพื่อแสดงโหนดของคู่คำที่มีความสัมพันธ์กัน ซึ่งได้ผลการวิเคราะห์ดังนี้

กลุ่มที่ 1 แสดงถึงคู่คำที่สัมพันธ์กับงานวิจัยบรรณมิติในแง่สถานการณ์ต่างๆ ในการกำหนดกรอบ และทิศทางในการวิจัย ดังคำที่พบ เช่น current: ปัจจุบัน, trends: แนวโน้ม, status: สถานะหรือสถานการณ์, productivity: ผลิตภาพหรือประสิทธิภาพในการผลิตผลงาน, global: ทั่วโลก, agenda: กำหนดการ, future: อนาคต, research: การวิจัย, emerging: กำลังเกิดขึ้น หรือเป็นการเจริญเติบโต, output: ผลผลิต, hotspots: จุดเด่น หรือพื้นที่ที่มีการเกิดเหตุการณ์หรือกิจกรรมที่น่าสนใจ, landscape: ทัศนียภาพ หรือสถานการณ์โดยรอบ เป็นต้น

กลุ่มที่ 2 แสดงถึงคู่คำที่สัมพันธ์กับงานวิจัย โดยการวัดปริมาณและคุณภาพของงานวิจัย การวิเคราะห์เชิงระบบ การทบทวนข้อมูล การเชื่อมโยงข้อมูลเพื่อแสดงผลในรูปแบบภาพ ซึ่งพบคำว่า literature: สิ่งพิมพ์/วรรณกรรม, systematic: อย่างเป็นระบบ, review: การทบทวน หรือการตรวจสอบข้อมูล, visualized: ทำให้เห็นภาพได้ชัดเจน เช่น ทำเป็นแผนภูมิหรือผังความสัมพันธ์, bibliometric: บรรณมิติ, network: ระบบเครือข่าย หรือการเชื่อมโยงข้อมูลแบบกราฟ, visual: ภาพหรือสิ่งที่เกี่ยวกับการมองเห็น, based: พื้นฐาน, approach: แนวทาง, overview: ภาพรวม, analysis: การวิเคราะห์, visualization: การแสดงแผนภาพ เป็นต้น

กลุ่มที่ 3 แสดงคู่คำที่สัมพันธ์กับงานวิจัยบรรณมิติที่มีการสืบค้นและวิเคราะห์บทความวิจัยที่ถูกอ้างอิงมากที่สุด โดยใช้วิธีการวัดผลด้วยตัวเลขของบทความที่ถูกอ้างอิงซึ่งเป็นตัวบ่งชี้ว่างานวิจัยนั้นๆ ได้รับความนิยมและมีผลกระทบในวงกว้างวิชาการในระดับสูง ได้แก่ top: อันดับสูงสุด, 100: จำนวน 100, cited: การอ้างอิง, papers: เอกสาร, highly: อย่างสูงหรืออย่างมาก เป็นต้น

นอกจากนี้ ภาพประกอบ 5 ยังมีคู่คำบางคู่ที่ไม่ได้เชื่อมโยงกับกลุ่มคำสามกลุ่มที่เหลือ ซึ่งเป็นตัวบ่งชี้ว่างานวิจัยด้านบรรณมิติที่เกี่ยวข้องในด้านอื่นๆ ด้วย เช่น supply-chain: ห่วงโซ่อุปทาน, covid-19-pandemic: โรคระบาดโควิด-19, intellectual-structure: โครงสร้างการเรียนรู้, Scopus-database: ฐานข้อมูล Scopus เป็นฐานข้อมูลที่รวบรวม จัดเก็บ ทำดัชนีผลงานทางวิชาการที่ใช้สำหรับการค้นคว้าผลงานวิจัย, mapping-knowledge-domain: แผนภาพความรู้ เพื่อแสดงแนวโน้มและการพัฒนาในสาขาวิชาการต่างๆ, information-science: สารสนเทศศาสตร์ เป็นการศึกษาวิธีการจัดการข้อมูลและเทคโนโลยีสารสนเทศ, social-media: สื่อสังคมออนไลน์ สื่อที่สำคัญในช่วยในการสื่อสารและแลกเปลี่ยนข้อมูลในเครือข่ายทางอินเทอร์เน็ต ในขณะเดียวกันพบคำที่เป็นตัวเลข คือ 2010 - 2019 อาจบ่งบอกถึงงานวิจัยที่เกี่ยวข้องกับบรรณมิติในช่วงเวลาปี.ศ. 2010 - 2019 ซึ่งใช้เป็นชื่อเรื่องในงานวิจัยด้านนี้มากที่สุด เป็นต้น

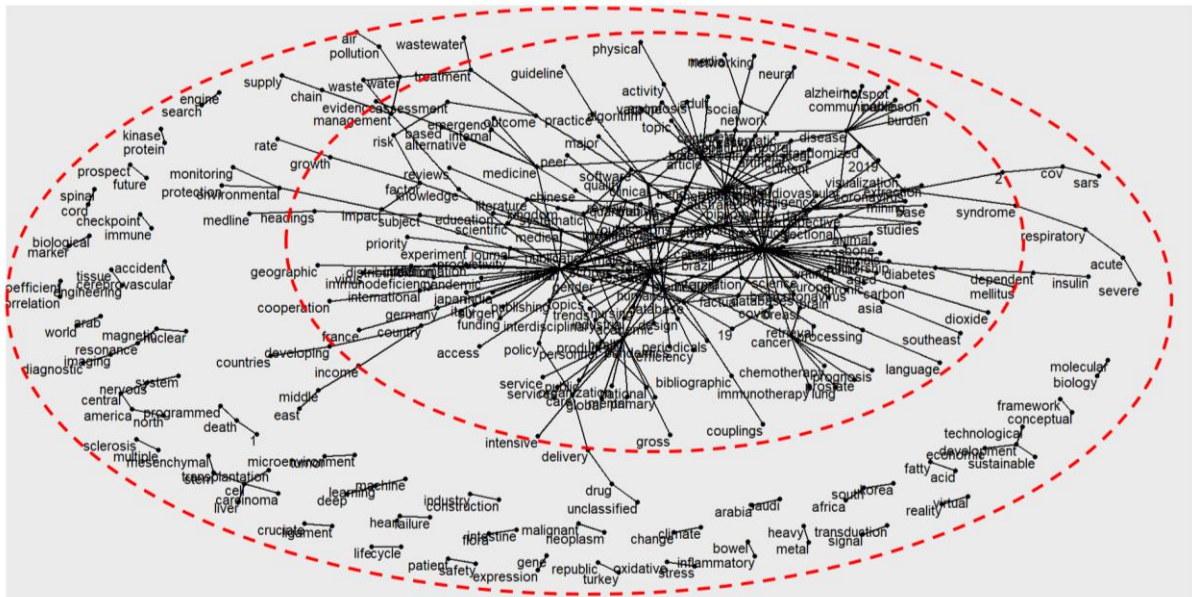


ภาพประกอบ 5 การวิเคราะห์ชื่อเรื่องในงานวิจัยด้านบรรณมิติ (โทเค็นคำ)

4. ผลการวิเคราะห์แผนภาพความสัมพันธ์ของโทเค็นคำ (word tokenization): ดัชนีคำสำคัญ

ภาพประกอบ 6 พบว่า ผลการวิเคราะห์โดยการทำโทเค็นคำดัชนีคำสำคัญ มีลักษณะความสัมพันธ์ที่ใกล้ชิดกันมากที่สุดเป็นกลุ่มขนาดใหญ่และมีความทับซ้อนของคำ เนื่องจากมีความถี่ในการใช้คำมากที่สุดในระดับสำคัญของงานวิจัยนั้นๆ โดยสังเกตจากวงกลมวงในสุด พบดัชนีคำสำคัญ ได้แก่ systematic: อย่างเป็นระบบ, medical: ทางกายภาพ, health: สุขภาพ, care: การดูแล, china: ประเทศจีน, impact: ผลกระทบ, factor: ปัจจัย, covid-19: โรคโควิด-19, biomedical: ชีวภาพ, การแพทย์, coronavirus: โควิดไวรัส, disease: โรค, artificial: สิ่งที่ถูกสร้างขึ้นมาจากการใช้เทคโนโลยีหรือคอมพิวเตอร์, kingdom: อาณาจักร, sustainable: ความยั่งยืน, geographic: ภูมิศาสตร์, machine: เครื่องจักร, clinical: ทางคลินิก, risk: ความเสี่ยง, climate: สภาพอากาศ, brazil: ประเทศบราซิล, policy: นโยบาย, carbon: คาร์บอน, asia: ทวีปเอเชีย, social: สังคม, service: การบริการ เป็นต้น ในขณะที่ส่วนวงกลมสีแดงวงนอกใน ภาพประกอบ 5 นั้นพบว่า มีดัชนีคำสำคัญที่เกี่ยวข้องที่น่าสนใจ เช่น air-pollution: มลพิษทางอากาศ, cancer-prognosis: การพยากรณ์โรคมะเร็ง, kinase protein: โปรตีนคินาส, wastewater-treatment: การบำบัดน้ำเสีย, life-cycle: วงจรชีวิต, sars-cov-2: ไวรัสโคโรนา, fatty-acid: กรดไขมัน, spinal-cord: สมองส่วนหลัง, immune-checkpoint: จุดควบคุมระบบภูมิคุ้มกัน, safety-patient: ความปลอดภัยของผู้ป่วย, biological-market: ตลาดชีวภาพ เป็นต้น

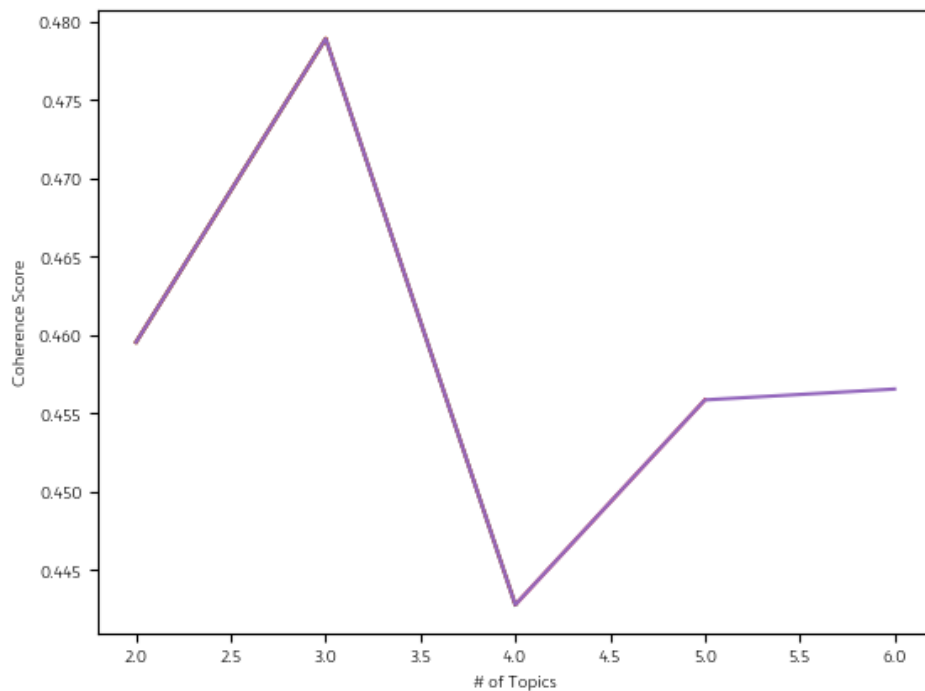
สรุปผลวิเคราะห์ข้อมูลใน ภาพประกอบ 6 ของการใช้ดัชนีคำสำคัญในฐานข้อมูล Scopus พบว่า มีงานวิจัยที่มีความเกี่ยวข้องกับด้านวิทยาศาสตร์สุขภาพมากที่สุด ซึ่งเป็นการยืนยันให้เห็นถึงความสำคัญของงานด้านนี้ที่เกี่ยวข้องกับสุขภาพและทางการแพทย์มากที่สุด นอกจากนี้ยังพบว่า งานวิจัยบรรณมิติพบในการศึกษาด้านบริหารจัดการ ด้านสังคมศาสตร์ และเทคโนโลยีสารสนเทศ ซึ่งมีส่วนร่วมสูงในการสร้างผลงานวิจัยในช่วงปี ค.ศ.2019-2023 นอกจากนี้ ภาพประกอบ 6 ยังแสดงให้เห็นถึงประเทศที่มีการทำวิจัยบรรณมิติสูงสุดซึ่งพบว่า มีประเทศจีน เป็นประเทศที่มีการทำงานวิจัยมากที่สุด รวมถึงประเทศบราซิล ญี่ปุ่น สหราชอาณาจักร และทวีปเอเชีย เป็นประเทศที่มีการศึกษาวิจัยด้านบรรณมิตินี้ด้วยเช่นกัน



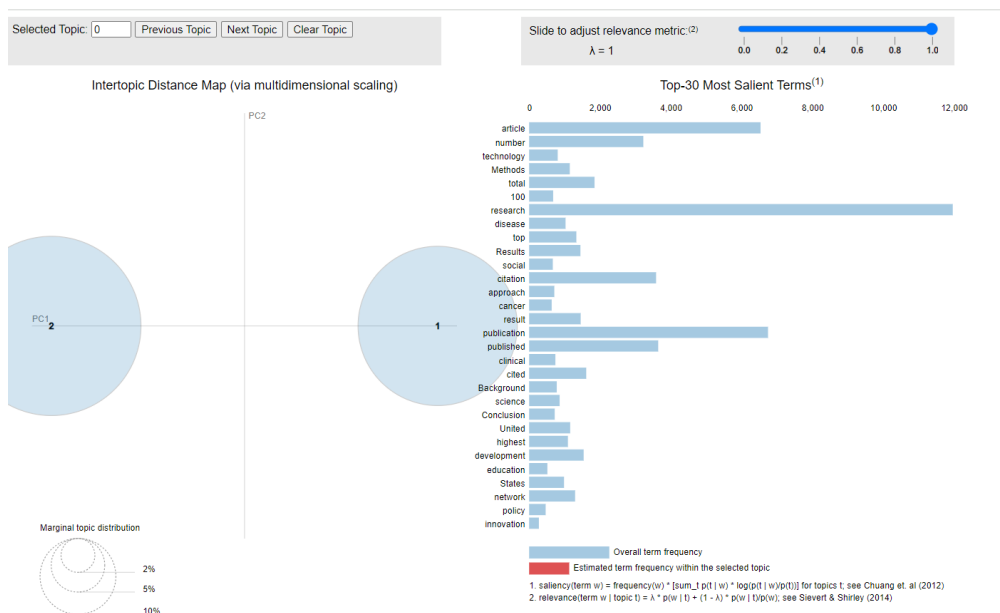
5. ผลการวิเคราะห์การแบ่งกลุ่ม Latent Dirichlet Allocation (LDA) และแผนภาพด้วย pyLDAvis

5. ผลการวิเคราะห์การแบ่งกลุ่ม Latent Dirichlet Allocation (LDA) และแผนภาพด้วย pyLDAvis

51



ภาพประกอบ 7 การวัดผลการแบ่งกลุ่ม LDA ด้วย Coherence Metrics Technique



ภาพประกอบ 8 ผลการวิเคราะห์แบบจำลองด้วย pyDAVis

https://www.huso-kku.org/bibliometric/lda_bibliometric_studies.html

อภิปรายผลการวิจัย

การศึกษาวิจัยครั้งนี้ใช้ระบบมือ (Manual system) ในการวิเคราะห์ข้อมูลด้วยโปรแกรม R studio และ Python ในการสร้างโค้ดและผลลัพธ์ด้วยแผนภาพเพื่อการวิเคราะห์ อย่างไรก็ตาม ผู้วิจัยใช้ประยุกต์ใช้ภาษา R และ Python เพื่อการวิเคราะห์ข้อมูลในการวิเคราะห์แบบจำลองหัวข้อด้วยเทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) ซึ่งโปรแกรม

R และ Python วิเคราะห์และสร้างแผนภาพความสัมพันธ์แทนแนวความคิดด้วยเทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) ทำให้เห็นแบบจำลองตามหัวข้อของงานวิจัยด้านบรรณมิติตามวัตถุประสงค์ของงานวิจัยที่กำหนด ในขณะเดียวกัน งานวิจัยของ Choi and Um (2023) เน้นการใช้เทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) เพื่อวิเคราะห์บทความข่าวที่เกี่ยวข้องกับ Covid-19 สำหรับการระบุหัวข้อที่สำคัญของบทความข่าวในช่วงการระบาดของโรค ซึ่งมีคำสำคัญที่พบมากที่สุดในการศึกษา ได้แก่ คำว่า Patients: ผู้ป่วย, New types : ชนิดใหม่, Infections: การติดเชื้อ, Symptoms: อาการ, Confirmation: การยืนยัน, ผู้ป่วยที่ได้รับการยืนยัน, Death: การเสียชีวิต, Occurrence: การเกิดขึ้น, Day: วัน, Investigation: การสอบสวน, Addition: การเพิ่มขึ้น เป็นต้น อีกทั้ง สอดคล้องกับงานวิจัยของ Bedekar and Desai (2022) ที่วิเคราะห์โดยใช้ชื่อเรื่องของงานวิจัยที่มีค่าสำคัญคำว่า ‘Covid-19’ โดยใช้เทคนิคการสร้างแผนภาพ ซึ่งจะเห็นได้ว่า มีงานวิจัยที่ใช้เทคนิคโดยใช้ Keywords ในการศึกษาด้วย นอกจากนี้ Qi et al. (2020) ศึกษาวิจัยโดยอาศัยเทคนิคการวิเคราะห์โทเค็นคำ (word token) ด้วยภาษา Python อาจมีความคล้ายคลึงกับการศึกษาในครั้งนี้ที่มีส่วนการวิเคราะห์ word token ดังที่แสดงผลในภาพประกอบ 5 - 6 ด้วย ในขณะเดียวกัน Crossley et al. (2014) วิจัยผลกระทบจากความถี่และการเรียนรู้คำศัพท์ของภาษาที่สองด้วย word type, word token และ word production ส่วน Kwok et al. (2021) ศึกษาการแสดงผลด้วยแผนภาพ word cloud ที่มีความถี่สูงและความสัมพันธ์ระหว่าง word tokens และสร้างแบบจำลองหัวข้อด้วย Latent Dirichlet Allocation (LDA) เพื่อระบุหัวข้อที่พูดถึงบ่อยจากผู้ใช้ Twitter ที่เกี่ยวข้องกับคำสำคัญ ‘วัคซีน COVID-19’ ในปี ค.ศ. 2020 ประเทศออสเตรเลีย เป็นต้น อย่างไรก็ตาม เทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) ในการวิเคราะห์บทความวิจัยในหลายสาขาวิชา อาจจะต้องพิจารณาความเหมาะสมกับหัวข้อและวัตถุประสงค์ของงานวิจัยด้วย โดยใช้เทคนิคที่เหมาะสมเพื่อให้ได้ผลลัพธ์ที่แม่นยำและมีประสิทธิภาพ ซึ่งการใช้เทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) ในการศึกษาครั้งนี้สามารถวิเคราะห์หัวข้อหรือคำสำคัญของงานวิจัยได้ โดยพบคำที่ใช้บ่อยในบทความวิจัยเป็นตัวชี้วัด อย่างไรก็ตาม การตีความผลลัพธ์หรือการแปลผลที่ได้จากการสร้างแบบจำลองตามหัวข้อ (Topic modeling) อาจต้องพิจารณาคำที่เลือกมาเป็นคำสำคัญว่าตรงกับหัวข้อของบทความนั้นจริงหรือไม่ ซึ่งงานวิจัยนี้ได้นำเสนอการวิเคราะห์ชื่อเรื่องและดัชนีคำสำคัญของงานวิจัยด้านบรรณมิติโดยใช้เทคนิคการสร้างแบบจำลองตามหัวข้อ โดยทดสอบประสิทธิภาพของเทคนิคการสร้างแบบจำลองตามหัวข้อ (Topic modeling) พบว่า คำที่พบสูงสุดในชื่อเรื่องของงานวิจัยดังกล่าว ส่วนใหญ่เกี่ยวข้องกับวิทยาศาสตร์สุขภาพและทางการแพทย์ ซึ่งสอดคล้องกับดัชนีคำสำคัญที่มีผลการวิเคราะห์ที่คล้ายและไม่แตกต่างกัน นอกจากนี้ยังมีงานวิจัยด้านบรรณมิติที่เกี่ยวข้องกับการศึกษา การบริหารจัดการ สังคมศาสตร์ เทคโนโลยีสารสนเทศ และสิ่งแวดล้อม ซึ่งการศึกษานี้ช่วยให้เห็นสภาพแบบจำลองตามหัวข้อในการวิเคราะห์ข้อมูลงานวิจัยด้านบรรณมิติในช่วงระหว่างปี ค.ศ. 2019-2023 ซึ่งมีประโยชน์ต่อผู้ที่ต้องการวิเคราะห์งานวิจัยในอนาคต โดยเฉพาะอย่างยิ่งกับนักวิชาการหรือนักวิจัยที่สนใจการทำงานวิจัยด้านบรรณมิติและการออกแบบจำลองหัวข้อได้

ข้อเสนอแนะ

1. เพื่อให้เห็นถึงแบบจำลองตามหัวข้อของงานวิจัยด้านบรรณมิติสำหรับการต่อยอดและพัฒนาหัวข้องานวิจัยบรรณมิติในสาขาวิชาหรือองค์ความรู้ที่เป็นข้อมูลขนาดใหญ่ (big data) ด้านอื่นๆ ได้
2. ประยุกต์กับโปรแกรมอื่นที่มีประสิทธิภาพในการวิเคราะห์ข้อมูลด้วยเทคนิคที่แตกต่างกัน เช่น โปรแกรม Bibliometrix-R package, VOSviewer, CiteSpace หรือวิเคราะห์เอกสาร/งานวิจัยจากฐานข้อมูลอื่นๆ เช่น Web of Science, EBSCO และ ProQuest เป็นต้น

อ้างอิง

- Agarwal, A., Durairajanayagam, D., Tatagari, S., Esteves, S. C., Harlev, A., Henkel, R., Roychoudhury, S., Homa, S., Puchalt, N. G., Ramasamy, R., Majzoub, A., Ly, K. D., Tvrda, E., Assidi, M., Kesari, K., Sharma, R., Banihani, S., Ko, E., Abu-Elmagd, M., Gosalvez, J., & Bashir, A. (2016). Bibliometrics: tracking research impact by selecting the appropriate metrics. *Asian Journal of Andrology*, 18(2), 296-309.
- Bedekar, M., & Desai, S. (2022). Analysis of research paper titles containing covid-19 keyword using various visualization techniques. *Smart Innovation, Systems and Technologies*, 302, 115–122.
- Chander, R., Dhar, M., & Bhatt, K. (2022). Bibliometric analysis of studies on library security issues in academic Institutions. *Journal of Access Services*, 19, 86-104.
- Choi, Y. - J., & Um, Y. - J. (2023). Topic models to analyze disaster-related newspaper articles: focusing on COVID-19. *International Journal of Mental Health Promotion*, 25(3), 421-431.
- Crossley, S., Salsbury, T., Titak, A., & McNamara, D. (2014). Frequency effects and second language lexical acquisition word types, word tokens, and word production. *International Journal of Corpus Linguistics*, 19(3), 301-332.
- Deshmukh, D., Patil, R., Chafale, S., & Basutiya, S. (2019). Twitter sentiment analysis using R-studio. *Journal of Emerging Technologies and Innovative Research*, 6(5), 7-12.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296.
- Ellegaard, O., & Wallin, J.A. (2015). The bibliometric analysis of scholarly production: How great is the impact?. *Scientometrics*, 105, 1809–1831.
- Garcia-Zorita, C., & Pacios, A. R. (2018). Topic modelling characterization of Mudejar art based on document titles. *Digital Scholarship in the Humanities*, 33(3), 529–539.
- Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1-30.
- Grün, B., & Hornik, K. (2023, April 14). *Topicmodels: An R package for fitting topic models*. <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management*, 59, 467–483.
- Isoaho, K., Gritsenko, D., & Makela, E. (2021). Topic modeling and text analysis for qualitative policy research. *Policy Studies Journal*, 49(1), 300-324.
- Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: Machine learning analysis. *Journal of Medical Internet Research*, 23(5), e26953.
- Liu, X., Zhang, L., & Hong, S. (2011). Global biodiversity research during 1900–2009: A bibliometric analysis. *Biodiversity and Conservation*, 20(4), 807-826.

- Natukunda, A., & Muchene, L. K. (2023). Unsupervised title and abstract screening for systematic review: a retrospective case-study using topic modelling methodology. *Systematic Reviews*, 12(1), 1-16.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, 25(4), 348-349.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020, July 5-10). *Stanza: A Python natural language processing toolkit for many human languages*. [Conference session]. The 58th Annual Meeting of the Association for Computational Linguistics, Seattle, Washington, USA.
- Saxton, M. D. (2018). A gentle introduction to topic modeling using python. *Theological Librarianship*, 11(1), 18-27.
- Sievert, C. & Shirley, K. (2014, June). *LDAvis: A method for visualizing and interpreting topics*. In Proceedings of the workshop on interactive language learning, visualization, and interfaces. (pp. 63-70).
- Sweileh, W. M. (2020). Bibliometric analysis of peer-reviewed literature on climate change and human health with an emphasis on infectious diseases. *Globalization and Health*, 16(1), 44.
- Wu, Y. - C., Chen, C. - S., & Chan, Y. - J. (2020). The outbreak of COVID-19: An overview. *Journal of the Chinese Medical Association*, 83(3), 217–220.
- Yeo Jin, J., & Youngmin, K. (2023). Research trends of sustainability and marketing research, 2010–2020: Topic modeling analysis. *Heliyon*, 9(3), e14208.
- Zhang, S., Ly, L., Mach, N., & Amaya, C. (2022). Topic modeling and sentiment analysis of yelp restaurant reviews. *International Journal of Information Systems in the Service Sector*, 14(1), Article no.72.
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472.