

Detecting Latent Topics for Categorizing Educational Dissertations of Leading Universities in the United States and Thailand: Structural Topic Modeling Analysis

การวิเคราะห์หัวข้อเรื่องแฝงเพื่อจัดหมวดหมู่หัวข้อวิทยานิพนธ์ทางการศึกษา
ของมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาและประเทศไทย: การวิเคราะห์แบบจำลองหัวข้อ

Chayut Piromsombat, Thanapa Ngewtong*, Kanessha Sirisak,
Saranunt Anuchon, and Teerayut Satjabut

ชยุตม์ ภิรมย์สมบัติ, ธนาภา จีวทอง*, กณิสสา ศิริศักดิ์, สรณันท์ อนุชน, และ ธีรยุทธ สัจจะบุตร

Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

*Corresponding author: thanapaani@gmail.com

Received August 11, 2023 ■ Revised October 31, 2023 ■ Accepted November 7, 2023 ■ Published December 28, 2023

Abstract

This research aimed to conduct a rigorous examination and comparison of latent topics of educational dissertations in leading universities in Thailand and the U.S. The dataset was acquired through web scraping, encompassing titles and abstracts of educational dissertations spanning a five-year period from 2019 to 2023. Specifically, we gathered 435 subjects in Thai from the Chulalongkorn University Institutional Repository (CUIR) and 363 subjects in English from the databases of the top 10 U.S. universities. The dataset was meticulously curated and processed utilizing Python's Natural Language Toolkit (NLTK) library. Subsequently, we performed an in-depth analysis employing Topic Modeling Analysis, with a focus on the Latent Dirichlet Allocation (LDA) method, BERTopic library, and Uniform Manifold Approximation and Projection (UMAP) model. The findings of this research underscore significant disparities in the topics explored within educational dissertations between Thailand and top-tier U.S. universities. In the context of Thailand, dissertation topics predominantly revolved around the development of teaching methodologies, student learning enhancement, and the resolution of teacher development issues. In contrast, dissertations from U.S. institutions encompassed a broader spectrum of themes, including area-based education, the development of ethnic and racial identity in students, and strategies for promoting equality among learners with diverse backgrounds. These outcomes highlight noteworthy variations in the academic emphasis and research priorities of these two distinct educational landscapes.

Keywords: topic modeling, educational dissertation, Latent Dirichlet Allocation (LDA), educational research, machine learning

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบหัวข้อเรื่องแฝง (Latent topics) ของวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาด้านการศึกษา ผู้วิจัยเก็บรวบรวมข้อมูลโดยใช้การดึงข้อมูลชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ทางการศึกษาจากเว็บไซต์ (Web scraping) บนฐานข้อมูลหลักของมหาวิทยาลัยในช่วงระยะเวลา 5 ปี (ค.ศ. 2019-2023) ได้จำนวนวิทยานิพนธ์ภาษาไทย 435 เรื่อง จากฐานข้อมูลคลังปัญญาจุฬาลงกรณ์ และภาษาอังกฤษ 363 เรื่อง จากฐานข้อมูลมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา ผู้วิจัยจัดเตรียมคลังข้อมูลให้อยู่ในรูปของรากศัพท์โดยใช้ไลบรารี Natural Language Toolkit ของ Python และวิเคราะห์ข้อมูลเชิงลึกโดยใช้การวิเคราะห์แบบจำลองหัวข้อ (Topic modeling analysis) ด้วยวิธีการจัดสรรดีริชเลตแฝง (Latent dirichlet analysis) โดยใช้ไลบรารี BERTopic และ Uniform Manifold Approximation and Projection (UMAP) model ผลการวิจัยพบว่า หมวดหมู่วิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกามีความแตกต่างกัน วิทยานิพนธ์ทางการศึกษาในประเทศไทยส่วนใหญ่เป็นการวิจัยด้านการพัฒนาการจัดการเรียนการสอน การส่งเสริมและพัฒนาการเรียนรู้ของผู้เรียน รวมถึงให้ความสำคัญในประเด็นวิจัยด้านการพัฒนาครูผู้สอน ในขณะที่หมวดหมู่วิทยานิพนธ์ทางการศึกษาของมหาวิทยาลัยชั้นนำในสหรัฐอเมริกามีประเด็นการวิจัยในลักษณะที่หลากหลายมากกว่า โดยเป็นการวิจัยในประเด็นที่เกี่ยวข้องกับการจัดการศึกษาในเชิงพื้นที่การพัฒนาผู้เรียนตามอัตลักษณ์และชาติพันธุ์ ตลอดจนการวิจัยที่มุ่งการจัดการเรียนการสอนให้เกิดความเท่าเทียมบนพื้นฐานความแตกต่างของผู้เรียน

คำสำคัญ: แบบจำลองหัวข้อ, วิทยานิพนธ์สาขาการศึกษา, การจัดสรรดีริชเลตแฝง, การวิจัยทางการศึกษา, การเรียนรู้ของเครื่อง

■ บทนำ (Introduction)

ปัจจุบันงานวิทยานิพนธ์ทางการศึกษาได้เกิดขึ้นเป็นจำนวนมากโดยเป็นการศึกษาค้นคว้าเกี่ยวกับศาสตร์ทางด้านการศึกษาซึ่งเกี่ยวกับความรู้ในด้านการเรียนการสอนที่ผู้สอนสามารถนำไปประยุกต์ใช้เพื่อทำให้ผู้เรียนได้เรียนรู้ตามเป้าหมายหรือวัตถุประสงค์ที่ตั้งไว้ ซึ่งประกอบไปด้วยปรัชญาการศึกษา บริบททางการสอน ทฤษฎี หลักการ แนวคิด ระบบ รูปแบบ วิธีการ เทคนิค และจิตวิทยาทางการเรียนรู้และการสอน การวางแผนและออกแบบการจัดการเรียนการสอน การดำเนินการเรียนการสอน การวัดและประเมินผล สื่อเทคโนโลยี ทางการสอน นวัตกรรมและการวิจัยการเรียนการสอน (Kaemane, 2010) ในช่วงหลายปีที่ผ่านมา พบว่าการวิจัยทางด้านการศึกษาในช่วงศตวรรษที่ 21 นักวิจัยยุคใหม่ได้ให้ความสนใจในงานวิจัยที่ใช้เทคโนโลยีเป็นพื้นฐานในการศึกษามากขึ้น ในทางกลับกัน ยังคงมุ่งเน้นการวิจัยที่อาศัยการปฏิบัติและแก้ไขปัญหาที่เกิดขึ้นจริง เช่น การวิจัยเชิงปฏิบัติการ การวิจัยประยุกต์ มากกว่าการศึกษาวิจัยเชิงทฤษฎี (Bhattacharya et al., 2021) สำหรับการวิจัยทางการศึกษาในช่วงหลายปีที่ผ่านมามีจำนวนมากขึ้นเรื่อย ๆ และมีการศึกษาในประเด็นที่หลากหลายตามสภาพปัญหาที่เกิดขึ้นในบริบทต่าง ๆ ความสนใจของผู้วิจัย นโยบายทางศึกษาของรัฐ หรือลักษณะสภาพเศรษฐกิจและสังคม ตลอดจนเทคโนโลยีที่เปลี่ยนแปลงไปอย่างรวดเร็ว

จากปริมาณวิทยานิพนธ์ทางการศึกษาที่มีอยู่เป็นจำนวนมากนั้น ล้วนมีคุณค่าและความสำคัญต่อการพัฒนาการศึกษาของประเทศเป็นอย่างมาก หากได้นำข้อค้นพบของวิทยานิพนธ์เหล่านี้ไปประยุกต์ใช้ให้เกิดประโยชน์เนื่องจากงานในกลุ่มนี้จะช่วยให้ได้องค์ความรู้และสารสนเทศที่เกี่ยวข้องกับปรัชญาการศึกษา การจัดการเรียนการสอนในรูปแบบต่าง ๆ จิตวิทยาทางการศึกษา การวัดและประเมินผล สื่อเทคโนโลยีทางการสอน ตลอดจนนวัตกรรมและการวิจัยการเรียนการสอน โดยข้อความรู้เหล่านี้ล้วนเป็นประโยชน์อย่างยิ่งต่อบุคลากรทางการศึกษาทั้งในระดับนโยบายและระดับปฏิบัติ ในระดับนโยบาย หากนำสารสนเทศที่ได้ไปประยุกต์ใช้จะช่วยให้การวางแผนนโยบายที่เกี่ยวข้องกับการศึกษาตอบสนองต่อความต้องการจำเป็นและเกิดประโยชน์สูงสุดต่อผู้สอนและผู้เรียน ในขณะเดียวกันประโยชน์ต่อครูในฐานะผู้ปฏิบัติก็สามารถนำความรู้ที่ได้จากข้อค้นพบไปประยุกต์ใช้ในการจัดการเรียนการสอนของตนให้มีประสิทธิภาพมากยิ่งขึ้น และช่วยให้พัฒนาความรู้ของผู้เรียนในด้านต่าง ๆ ได้อย่างมีคุณภาพ ดังนั้น การทำความเข้าใจขอบเขต แนวโน้ม และอนาคตของงานวิจัยทางการศึกษาเป็นสิ่งที่จำเป็นสำหรับบุคลากรทางการศึกษา นักวิชาการ นักวิจัยและผู้ที่มีส่วนเกี่ยวข้องในการทำความเข้าใจถึงลักษณะของหมวดหมู่ หรือหัวข้อที่มีการศึกษาในปัจจุบัน ตลอดจนแนวโน้มของการศึกษา

เพื่อนำไปต่อยอดในการศึกษาค้นคว้า หรือพัฒนาความรู้ตลอดจนพัฒนานวัตกรรมที่เกี่ยวข้อง รวมถึงช่วยให้เห็นถึงช่องว่างในการศึกษาซึ่งนำไปสู่การเติมเต็มความรู้ที่ยังขาดการศึกษาต่อไป

จากความสำคัญของการทำความเข้าใจถึงลักษณะของหมวดหมู่ หรือหัวข้อวิทยานิพนธ์ทางการศึกษาในปัจจุบัน นักวิจัยและนักวิชาการทางการศึกษาจึงได้ศึกษาวิเคราะห์ และจัดหมวดหมู่ประเด็นหรือหัวข้อที่มีการศึกษาวิจัยในปัจจุบัน ตลอดจนแนวโน้มของการศึกษาในช่วงเวลาต่าง ๆ เพื่อนำสารสนเทศดังกล่าวไปพัฒนา และต่อยอดองค์ความรู้ทางการศึกษาให้เกิดประโยชน์ดังที่กล่าวไปแล้วข้างต้น แต่อย่างไรก็ตามการวิเคราะห์และจัดหมวดหมู่วิทยานิพนธ์ทางการศึกษาที่มีปริมาณมากในฐานข้อมูล การวิเคราะห์ด้วยวิธีการแบบดั้งเดิมจึงเป็นเรื่องที่ทำได้ยาก เช่น การวิเคราะห์เนื้อหา (Content analysis) ซึ่งเป็นกระบวนการวิเคราะห์ข้อมูลเชิงคุณภาพรูปแบบหนึ่งที่มีถูกนำมาใช้ในการจัดหมวดหมู่ของข้อมูล แต่เนื่องจากการจัดหมวดหมู่ รูปแบบ และความสัมพันธ์ของข้อมูลด้วยการวิเคราะห์เนื้อหา เป็นการวิเคราะห์ที่ต้องอาศัยการพิจารณาหรือดุลยพินิจของนักวิจัยเป็นสำคัญ ความรู้และความชำนาญของผู้วิจัยจึงมีผลต่อการจัดหมวดหมู่ของข้อมูลด้วย (Aureli, 2017) ดังนั้น หากข้อมูลมีปริมาณมากเกินไป หรือผู้วิจัยมีความรู้ความชำนาญในเรื่องดังกล่าวไม่เพียงพออาจส่งผลต่อคุณภาพการวิเคราะห์ข้อมูลได้ ซึ่งวิธีการทางสถิติวิธีหนึ่งที่เข้ามาช่วยในการศึกษางานวิจัยจำนวนมากเพื่อทำความเข้าใจวิเคราะห์ และจัดหมวดหมู่หัวข้อ หรือประเด็นในงานวิจัย คือการวิเคราะห์แบบจำลองหัวข้อ หรือ Topic modeling (Ekin et al., 2023)

การวิเคราะห์แบบจำลองหัวข้อ เป็นการวิเคราะห์และจัดหมวดหมู่หัวข้อด้วยการวิเคราะห์เชิงปริมาณเป็นสำคัญ ข้อมูลต่าง ๆ จะถูกวิเคราะห์ หาความสัมพันธ์และจัดหมวดหมู่ผ่านโปรแกรมคอมพิวเตอร์ ทำให้บทบาทของผู้วิจัยลดลง และอาศัยการเรียนรู้ของเครื่อง (Machine learning) ในการทำงาน โดยการวิเคราะห์แบบจำลองหัวข้อเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) ซึ่งเป็นการวิเคราะห์ข้อมูลผ่านตัวอักษร เพื่อสร้างแบบจำลองการกระจายตัวของคำและนำมาใช้ในการจัดกลุ่มคำเพื่อค้นหา “หัวข้อ” ซึ่งเป็นชุดของคำที่มีความสัมพันธ์และปรากฏขึ้นด้วยกัน (Ekin et al., 2023; Schmiedel et al., 2019) หรืออาจกล่าวได้ว่า การวิเคราะห์แบบจำลองหัวข้อมุ่งเน้นไปที่การจัดกลุ่มของข้อความในเอกสาร ซึ่งในแต่ละเอกสารจะมีหัวข้อแฝงซ่อนอยู่ (Mazumder & Barui, 2021) การวิเคราะห์แบบจำลองหัวข้อจะวิเคราะห์เอกสารเพื่อระบุหัวข้อและจัดเตรียมสร้างกลุ่มคำ (Cluster) ที่เพียงพอตัวอย่างเช่น อัลกอริทึมการวิเคราะห์แบบจำลองหัวข้อสามารถระบุได้ว่า เอกสารขาเข้าเป็นสัญญา ใบแจ้งหนี้ ข้อร้องเรียน

หรืออื่น ๆ โดยพิจารณาจากเนื้อหาของเอกสาร การวิเคราะห์ความหมายแฝง (Latent Semantic analysis) และการจัดสรร Dirichlet แฝง (Latent Dirichlet analysis) เป็นวิธีการสร้างแบบจำลองหัวข้อหลักสองวิธีที่นิยมนำมาใช้ในวิเคราะห์ข้อมูลในลักษณะข้อความขนาดใหญ่เพื่อจัดหมวดหมู่หัวข้อ โดยให้ข้อมูลเชิงลึกที่เป็นประโยชน์และสนับสนุนการตัดสินใจที่ดีขึ้น (Sheridan, 2022) นอกจากนี้ การวิเคราะห์ดังกล่าวสามารถวิเคราะห์ข้อมูลได้หลากหลายรูปแบบ ตัวอย่างเช่น ความคิดเห็นบนสื่อสังคมออนไลน์ (Hong & Davison, 2010) หรือข้อมูลในลักษณะของบทความ (Walker et al., 2019) ดังนั้น การวิเคราะห์แบบจำลองหัวข้อจึงเป็นวิธีการวิเคราะห์ที่เหมาะสมกับข้อมูลปริมาณมากและไม่มีข้อจำกัดในด้านความซับซ้อนของข้อมูล หรือความยากในการตีความข้อมูลที่ไม่รู้จัก

การวิเคราะห์แบบจำลองหัวข้อได้ถูกนำมาประยุกต์ใช้ในหลากหลายด้านไม่ว่าจะเป็นด้านการบริหารธุรกิจ การตลาด การแพทย์ และอื่น ๆ ในทางการศึกษาที่เช่นกัน ซึ่งพบว่าถูกนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลทางการศึกษาในหลากหลายประเด็น ตัวอย่างเช่น การวิเคราะห์แนวโน้มของการวิจัยในสาขาฟิสิกส์ศึกษาเพื่อระบุสถานะของการวิจัยฟิสิกส์ศึกษาและช่วยเหลือนักวิจัยในการศึกษาในอนาคต จากการรวบรวมบทความใน American Journal of Physics (AJP) และ Physics Review Physics Education Research (PRPER) (Yun, 2020) การนำการวิเคราะห์แบบจำลองหัวข้อมาใช้ในการค้นหาปัญหาที่เกี่ยวข้องกับอีเลิร์นนิ่งในช่วงการแพร่ระบาดของสถานการณ์โควิด-19 จากความคิดเห็นบน Twitter (Mujahid et al., 2021) หรือการวิเคราะห์วิวัฒนาการของแนวโน้มการวิจัย คำที่ใช้บ่อยหัวข้อการวิจัยแฝง และแนวโน้มการวิจัยสำหรับหัวข้อเฉพาะทางด้านการใช้เทคโนโลยีในการจัดการเรียนการสอนคณิตศาสตร์จากบทความของบทความจาก Web of Science, ERIC, และ PsycINFO (Hwang et al., 2023) จึงเห็นได้ว่า การวิเคราะห์แบบจำลองหัวข้อสามารถนำมาประยุกต์ใช้ในทางการศึกษาได้หลากหลายรูปแบบทั้งการวิเคราะห์หัวข้อทางการวิจัยทางการศึกษาในหลากหลายประเด็น รวมถึงการค้นหาปัญหาหรือข้อคิดเห็นที่เกี่ยวข้องกับการศึกษาบนสื่อออนไลน์ ซึ่งทำให้ผู้ที่มีส่วนเกี่ยวข้องสามารถนำข้อมูลที่ได้ไปใช้ในการแก้ไขปัญหาพัฒนาการจัดการเรียนการสอน หรือพัฒนาหัวข้อในการดำเนินการวิจัยทางการศึกษาเพื่อเติมเต็มช่องว่างหรือต่อยอดความรู้ในอนาคตได้

สำหรับในประเทศไทย วิทยานิพนธ์ทางการศึกษานั้นได้ถูกจัดทำขึ้นเป็นจำนวนมากในช่วงหลายปีที่ผ่านมาเช่นกัน แต่อย่างไรก็ตาม ยังไม่ทราบถึงลักษณะของหัวข้อและแนวโน้มการศึกษาของวิทยานิพนธ์ในประเทศไทยว่า มีลักษณะหัวข้อหรือประเด็นในการศึกษาและแนวโน้มของหัวข้อวิทยานิพนธ์ในช่วงหลายปีที่ผ่านมาเป็นอย่างไร งานวิจัยนี้จึงมุ่งศึกษา

และวิเคราะห์ลักษณะหัวข้อและแนวโน้มของวิทยานิพนธ์ทางการศึกษาในประเทศไทย รวมทั้งมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา (U.S. News & World Report, 2023) ซึ่งเป็นเสมือนต้นแบบทางด้านการจัดการศึกษา โดยมุ่งเน้นการวิเคราะห์งานวิทยานิพนธ์ทางการศึกษาที่เกิดขึ้นในช่วงระยะเวลา 5 ปี ตั้งแต่ปี ค.ศ. 2019-2023 จากชุดข้อมูลชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ และนำมาเปรียบเทียบหัวข้อเรื่องแฝงหรือหมวดหมู่ของการทำวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา เพื่อทำความเข้าใจและวิเคราะห์ความคล้ายคลึงหรือความแตกต่างของหัวข้อเรื่องแฝงที่มีการศึกษาจากทั้งสองประเทศ ซึ่งจะช่วยให้ผู้ที่มีส่วนเกี่ยวข้องทางการศึกษาสามารถนำข้อความรู้ที่ได้ไปพัฒนาเพื่อเติมเต็มช่องว่างในการศึกษา รวมถึงค้นคว้าต่อยอดในประเด็นที่สนใจ ตลอดจนเป็นข้อมูลสำคัญที่ช่วยให้การทำวิทยานิพนธ์ทางการศึกษาของประเทศไทยพัฒนาทิศทางในการทำการวิจัยให้สอดคล้องหรือเติมเต็มประเด็นช่องว่างที่อาจจะยังไม่เคยมีการศึกษาในประเด็นดังกล่าว เมื่อเทียบกับลักษณะหัวข้อเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาของมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา

วัตถุประสงค์การวิจัย (Objectives)

1. เพื่อวิเคราะห์หัวข้อเรื่องแฝง (Latent topics) ของวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาในช่วงระยะเวลา 5 ปี (ค.ศ. 2019-2023)
2. เพื่อเปรียบเทียบลักษณะหัวข้อเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาในช่วงระยะเวลา 5 ปี (ค.ศ. 2019-2023)

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง (Theories and Related Literature)

แบบจำลองหัวข้อ (Structural Topic Modeling)

ในการศึกษาเพื่อให้เข้าใจเกี่ยวกับขอบเขต แนวโน้มของการศึกษาวิจัยในสาขาการศึกษาจากชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ในสาขาการศึกษาเพื่อสะท้อนการศึกษาวิจัยในห้วง 5 ปีที่ผ่านมาสำหรับวิทยานิพนธ์ในสาขาการศึกษาที่จะแสดงหัวข้อที่มีการศึกษาจำนวนมาก ลักษณะในการศึกษาวิจัย หรือช่องว่างสำหรับการศึกษาวิจัยนั้นต้องประยุกต์ใช้วิธีการวิเคราะห์จากวิทยาการข้อมูล ได้แก่ การวิเคราะห์เหมืองข้อความ (Text mining) ซึ่งเป็นเทคนิควิธีการประมวลผลเพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้น โดยแบบจำลองหัวข้อเป็นระเบียบวิธีการหนึ่งที่วิเคราะห์เพื่อให้ได้รูปแบบที่เกิดซ้ำหรือความเกี่ยวข้องเชื่อมโยงของคลังข้อความในชุดเอกสาร การสร้างแบบจำลองหัวข้อเป็นเครื่องมือสำหรับการวิเคราะห์ชุดข้อมูลเชิงคุณภาพจำนวนมาก

ด้วยวิธีที่ปรับขนาดได้และทำซ้ำได้

แบบจำลองหัวข้อเป็นเทคนิคการเรียนรู้ของเครื่อง (Machine learning) แบบไม่มีผู้สอน (Unsupervised learning) ใช้ทำการวิเคราะห์ข้อมูลตัวอักษร (Text) เพื่อสร้างแบบจำลองการกระจายตัวของคำเพื่อนำมาใช้ในการจัดกลุ่มคำเพื่อหา “หัวข้อ (Topic)” จากข้อมูลชุดข้อความของเอกสาร (Document) ทั้งหมด โดยหัวข้อเป็นชุดของคำ (Words) ที่มีความสัมพันธ์และปรากฏขึ้นด้วยกัน (Ekin et al., 2023; Schmiedel et al., 2019) ซึ่งมีพื้นฐานมาจากแนวคิดที่ว่า ข้อความในเอกสารเกิดจากการรวมตัวของหัวข้อต่าง ๆ ซึ่งแต่ละหัวข้อมีการแจกแจงความน่าจะเป็นของคำที่เกิดขึ้นหลาย ๆ คำ ในแต่ละหัวข้อนั้นทำให้สามารถทำความเข้าใจในชุดข้อมูลหรือชุดเอกสารได้มากขึ้นจากการศึกษาเหมืองข้อมูล (Data mining) โดยใช้วิธีการจัดสรรดีริชเลตแฝง (Latent dirichlet allocation - LDA) เพื่อหาความน่าจะเป็นจากคำที่ปรากฏในเอกสาร ซึ่งจะ เป็นหัวข้อแฝง (Latent topic) (Schmiedel et al., 2019)

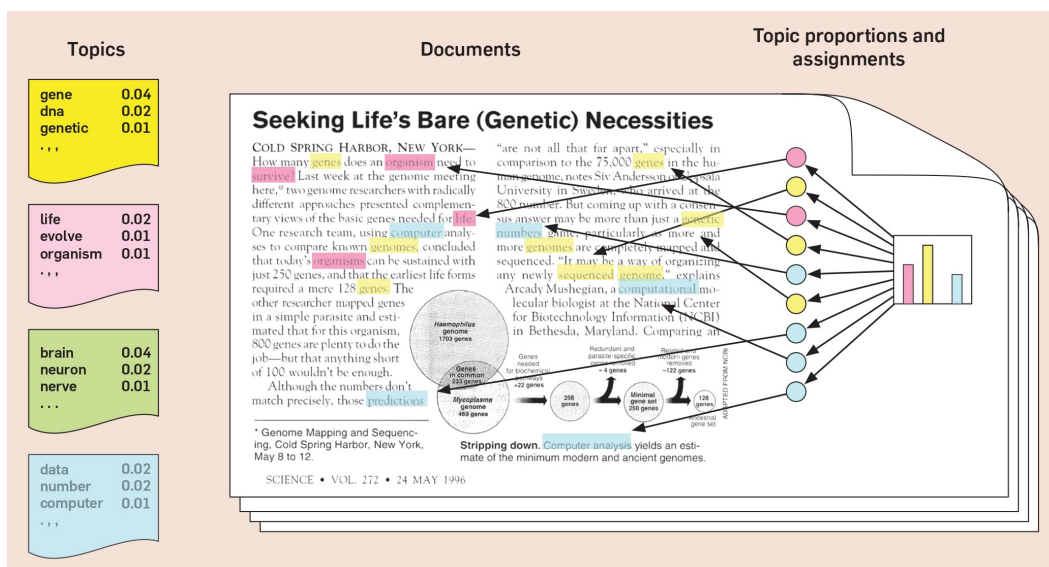
LDA เป็นกลุ่มของโมเดลที่เป็นไปได้โดยมีแนวคิดพื้นฐานคือ เอกสารประกอบด้วยหัวข้อแฝงที่ผสมกันหลากหลายอย่างสุ่ม แต่ละหัวข้อเกิดจากการกระจายของคำหลาย ๆ คำ (Blei et al., 2003; Chauhan & Shah, 2021) แบบจำลองความน่าจะเป็น สำหรับการทำแบบจำลองหัวข้อโดยหาหัวข้อที่ซ่อนอยู่ในเอกสาร LDA จะวิเคราะห์ความน่าจะเป็นของความสัมพันธ์

ของหัวข้อของแต่ละประโยค และค่าความน่าจะเป็นของคำใน แต่ละหัวข้อ ซึ่งใช้หลักความน่าจะเป็นในการหาพารามิเตอร์ (Probabilistic model) (Chauhan & Shah, 2021; Vayansky & Kumar, 2020) LDA เป็นวิธีที่ซับซ้อนที่สุด (Ekin et al., 2023) ที่มีความยืดหยุ่นและปรับเปลี่ยนได้สำหรับการสร้างแบบจำลอง หัวข้อ (Ekin et al., 2023; Vayansky & Kumar, 2020) โมเดล LDA มีกระบวนการทำงาน คือ 1) เลือกพารามิเตอร์โดยกำหนด จำนวนหัวข้อ (k) 2) LDA จะอ่านทุกคำในเอกสารและกำหนด หัวข้อให้แต่ละคำแบบสุ่ม ซึ่งจะผลลัพธ์เป็นกลุ่มคำของหัวข้อ (Topic representation) โดยจะมีค่าต่าง ๆ ที่อยู่ในหัวข้อนั้น ซึ่งคำที่อยู่ในแต่ละหัวข้อจะอยู่ในรูปแบบ Bag of words และ Topics ของแต่ละเอกสารโดยแสดงเป็นร้อยละความเป็นไปได้ ที่เอกสารจะอยู่ใน Topic ใดบ้าง โดย i) สุ่มเลือกหัวข้อจากการ แจกจ่ายหัวข้อในชั้น 1 และ ii) สุ่มเลือกคำจากการกระจาย หัวข้อของคำศัพท์ในเอกสารซึ่งเอกสารจะแสดงหลายหัวข้อ และแสดงหัวเรื่องแฝงในสัดส่วนที่ต่างกัน จาก Figure 1 แสดง กระบวนการทำงานเบื้องหลังของวิธีการจัดสรรดีริชเลตแฝง โดยมี “หัวข้อ” จำนวนหนึ่งซึ่งเป็นการแจกแจงของคำสำหรับ ชุดหัวเรื่องแฝงทั้งหมด (ซ้ายสุด) จากเอกสารแต่ละฉบับจะถูกสร้าง ในขั้นแรกจะเลือกการกระจายตามหัวข้อ (ฮิสโตแกรมทางด้านขวา) จากนั้นสำหรับแต่ละคำจะเลือกการกำหนดหัวข้อ (วงกลมสี) และ เลือกคำจากหัวข้อที่เกี่ยวข้อง (Blei, 2012)

Figure 1

The Intuitions Behind Latent Dirichlet Allocation (Blei, 2012)

วิธีการจัดสรรดีริชเลตแฝง (Latent Dirichlet Allocation-LDA) (Blei, 2012)



งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์แบบจำลองหัวข้อ จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการ วิเคราะห์ Topic modeling ในงานวิจัยพบว่า มีงานวิจัยหลาย

รูปแบบที่มีการใช้การวิเคราะห์ดังกล่าวเข้ามาประยุกต์ใน กระบวนการดำเนินการวิจัย โดยสามารถแบ่งตามแหล่งข้อมูล รูปแบบต่าง ๆ ที่นำมาใช้ในการวิเคราะห์ได้ดังตัวอย่างต่อไปนี้

1) ข้อความเกี่ยวกับการแสดงความคิดเห็นบนแพลตฟอร์มออนไลน์ เช่น การศึกษาของ Hong and Davison (2010) มีการศึกษาการใช้วิธีการ Topic modeling ในการวิเคราะห์ข้อความบนแพลตฟอร์ม Twitter (Steaming และ API) โดยใช้โมเดลหลักในการวิเคราะห์ 2 โมเดล ได้แก่ LDA และ Author-topic model มีวัตถุประสงค์เพื่อการทำนายข้อความที่ได้รับความนิยมสูงสุดบนแพลตฟอร์ม Twitter และเพื่อจัดประเภทผู้ใช้งาน Twitter และข้อความที่เกี่ยวข้องเป็นหมวดหมู่ตามประเด็น (Topic) ผลการวิจัยพบว่า การวิเคราะห์สามารถใช้ทำนายข้อความที่น่าจะได้รับความนิยมสูงสุดได้ โดยอาศัยข้อมูลจากการ Retweet และสามารถจัดหมวดหมู่ข้อมูลบนแพลตฟอร์ม Twitter ออกได้เป็น 16 หมวดหมู่ เช่น ศิลปะและการออกแบบ หนังสือ ธุรกิจ การกุศล ความบันเทิง สำหรับการศึกษาของ Amara et al. (2021) ที่ได้ทำการวิเคราะห์ Topic modeling จากข้อความที่เกี่ยวข้องกับการระบาดของเชื้อไวรัสโควิด-19 บนแพลตฟอร์มออนไลน์ Facebook โดยแบ่งการวิเคราะห์ออกเป็น 7 ภาษาที่มีการโพสต์หรือตอบโต้ (Comment) ได้แก่ ภาษาอังกฤษ อารบิก สเปน อิตาลี เยอรมัน ฝรั่งเศส และญี่ปุ่น เพื่อวิเคราะห์จัดหมวดหมู่ข้อความดังกล่าว

2) ข้อความในรูปแบบของข่าว (News) ในช่องทางต่าง ๆ โดยการศึกษาของ Al-Rawi et al. (2021) ทำการศึกษาการโพสต์

เกี่ยวกับข่าวจากช่องรายการต่าง ๆ บน Instagram เช่น CNN BBC France SkyNews เพื่อทำการจัดหมวดหมู่ข่าวที่มีการเผยแพร่ด้วยวิธีการ Topic modeling ผลการวิเคราะห์พบว่าสามารถแบ่งได้เป็น 7 หมวดหมู่ เช่น การเมือง กีฬา พิธีการ ความสนใจของคน ข่าวสารทั่วไป รวมถึงการวิเคราะห์ข่าวจากหนังสือพิมพ์ โดย Yang et al. (2011) ได้ทำการรวบรวมข่าวในหนังสือพิมพ์ของเมือง Texas ตั้งแต่ปี ค.ศ. 1829-2008 ทำการวิเคราะห์ Topic modeling จากเนื้อหาในข่าว โดยใช้โมเดล การวิเคราะห์รูปแบบ pLAS และ LDA ผลการวิเคราะห์พบว่า สามารถแบ่งเนื้อหาของข่าวในหนังสือพิมพ์ออกได้เป็น 10 หมวดหมู่

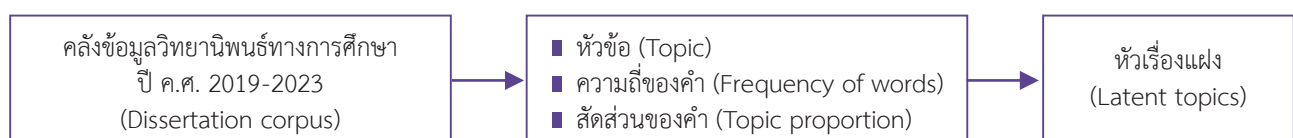
3) ข้อมูลในลักษณะของบทความ (Article) การศึกษาของ Walker et al. (2019) ได้ทำการวิเคราะห์ข้อมูลที่อยู่ในรูปของบทความจาก 2 วารสาร ได้แก่ PAR และ PAT ที่จัดพิมพ์โดย American Society for Public Administration (ASPA) จำนวน 3,796 บทความ วิเคราะห์ด้วยวิธีการทำ Topic modeling โดยใช้ชื่อเรื่องและบทคัดย่อเป็นข้อมูลในการวิเคราะห์ ผลการวิเคราะห์ทำให้สามารถจัดหมวดหมู่บทความจากทั้งสองวารสารออกได้เป็น 9 หมวดหมู่ ตัวอย่างเช่น การมีส่วนร่วมของประชาชน สุขภาวะ การจัดการทางการเงิน การเลือกตั้ง

■ กรอบแนวคิดการวิจัย (Conceptual framework)

Figure 2

Conceptual Framework

กรอบแนวคิดในการวิจัย



จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องผู้วิจัยมุ่งวิเคราะห์หัวข้อเรื่องแฝงจากคลังข้อมูลวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาในช่วงระยะเวลา 5 ปี (ค.ศ. 2019-2023) โดยใช้การวิเคราะห์แบบจำลองหัวข้อ ในการวิเคราะห์ข้อมูลตัวอักษรเพื่อสร้างแบบจำลองการกระจายตัวของคำเพื่อนำมาใช้ในการจัดกลุ่มคำจากข้อมูลชุดชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ในภาษาอังกฤษ นอกจากนี้ ผู้วิจัยยังมุ่งเปรียบเทียบลักษณะของหัวข้อเรื่องแฝงที่ได้จากการวิเคราะห์ระหว่างวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาเพื่อศึกษาช่องว่างในการศึกษา ตลอดจนเป็นข้อมูลสำคัญที่ช่วยให้การทำวิทยานิพนธ์ทางการศึกษาของประเทศไทยพัฒนาทิศทางในการทำวิจัยให้สอดคล้องหรือเพิ่มเติมประเด็นช่องว่างที่อาจจะยัง

ไม่เคยมีการศึกษาในประเด็นดังกล่าวต่อไปในอนาคตโดยอาศัยเทคนิคการวิเคราะห์ข้อมูลที่มีความได้เปรียบสำหรับชุดข้อมูลที่มีปริมาณมากด้วยเทคนิคการเรียนรู้ของเครื่อง แบบไม่มีผู้สอน แทนการวิเคราะห์แบบดั้งเดิม

■ ขอบเขตของการวิจัย (Scope)

การวิจัยเรื่องนี้เก็บรวบรวมข้อมูลจากวิทยานิพนธ์ด้านการศึกษาในช่วงระยะเวลา 5 ปี (ค.ศ. 2019-2023) ด้วยวิธีการดึงข้อมูลชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ในรูปแบบภาษาอังกฤษจากเว็บไซต์ในฐานข้อมูลของแต่ละกลุ่มมหาวิทยาลัย แบ่งออกเป็น 2 กลุ่มหลัก ได้แก่ 1) วิทยานิพนธ์ในประเทศไทยจากฐานข้อมูล CUIR ของคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยเพียงฐานข้อมูลเดียว เนื่องจาก

ได้รับการจัดอันดับจาก QS World University Rankings ปี ค.ศ. 2018-2021 และ Times Higher Education World University Rankings ปี 2020 รายสาขาด้านการศึกษา (Education) เพียงแห่งเดียวในประเทศไทยที่อยู่ในอันดับโลก (The QS World University Rankings, 2021; The Times Higher Education World University Rankings, 2020) จำนวน 435 เรื่อง และ 2) วิทยานิพนธ์ในประเทศสหรัฐอเมริกา จากมหาวิทยาลัยที่ติดอันดับ (Top 10) ด้านการศึกษา (General education) จากการจัดอันดับของ U.S. News & World Report (U.S. News & World Report, 2023) ซึ่งประกอบด้วย University of Pennsylvania, Harvard University, University of California--Los Angeles, New York University (Steinhardt), University of Wisconsin--Madison, Vanderbilt University (Peabody), Teachers College, Columbia University, University of Michigan--Ann Arbor, Stanford University และ University of California--Irvine แต่เนื่องจากข้อจำกัดในด้านการเข้าถึงฐานข้อมูลบนเว็บไซต์ของมหาวิทยาลัย ทำให้สามารถเข้าถึงแหล่งข้อมูลได้เพียง 3 มหาวิทยาลัย ประกอบด้วย Harvard University, UCLA-University of California--Los Angeles และ Vanderbilt University (Peabody) รวมทั้งหมด จำนวน 363 เรื่อง

■ วิธีดำเนินการวิจัย (Methodology)

1. กระบวนการรวบรวมข้อมูล (Data retrieval) ผู้วิจัยดำเนินการรวบรวมข้อมูลด้วยวิธีการดึงข้อมูลจากเว็บไซต์ (Web scraping) บนฐานข้อมูลวิทยานิพนธ์สาขาการศึกษา (General education) ของมหาวิทยาลัยในประเทศไทยและประเทศสหรัฐอเมริกาที่สามารถเข้าถึงได้และมีการเผยแพร่ระหว่างปี ค.ศ. 2019-2023 โดยดึงข้อมูลชื่อเรื่องและบทคัดย่อของวิทยานิพนธ์ในภาษาอังกฤษจัดทำเป็นฐานข้อมูล (Raw data)

2. กระบวนการเตรียมข้อมูล (Data preparation) หลังจากเก็บรวบรวมข้อมูลวิทยานิพนธ์จากทุกมหาวิทยาลัยแล้ว จึงนำข้อมูลที่ได้มาตรวจสอบและคัดกรองวิทยานิพนธ์ในสาขาการศึกษาและจัดรูปแบบข้อมูลให้อยู่ในรูปแบบที่เหมาะสมในการวิเคราะห์ ด้วยการจัดกระทำข้อมูลให้เป็นเอกสารมาตรฐานเดียวกัน (Document standardization) โดยประกอบด้วย ชื่อเรื่องและบทคัดย่อภาษาอังกฤษที่อยู่ในรูปแบบไฟล์ประเภท Flat file (CSV) เริ่มดำเนินการด้วยการลบข้อมูลบางส่วน ได้แก่ 1) การกำหนดตัวเลขที่ต้องการนำออก (Number) 2) การกำหนดเครื่องหมายวรรคตอน (Punctuation) 3) การกำหนดแถบช่องว่างระหว่างคำ (Strip whitespace) และ 4) การจัดกระทำข้อมูลที่มีความหมายเหมือนกันให้เป็นข้อมูลเดียวกัน ซึ่งเป็นกระบวนการเปลี่ยนคำให้อยู่ในรูปของรากศัพท์ (Root form) โดยใช้

Natural Language Toolkit (Nltk) ที่เป็นไลบรารีของไพทอน ซึ่งเป็นชุดโปรแกรมสำหรับการประมวลผลภาษาธรรมชาติเชิงสถิติสำหรับภาษาอังกฤษ เพื่อช่วยในการเตรียมข้อมูลก่อนที่จะนำข้อมูลประเภทข้อความไปให้คอมพิวเตอร์ทำการประมวลผล

3. การสร้างแบบจำลองหัวข้อ (Topic modeling) ในขั้นตอนนี้ผู้วิจัยใช้ชุดข้อมูลที่ผ่านกระบวนการเตรียมข้อมูลทำการวิเคราะห์แบบจำลองหัวข้อด้วยวิธี LDA โดยใช้ไลบรารี BERTopic ในโปรแกรมไพทอน โดยกำหนด UMAP (Uniform Manifold Approximation and Projection) โมเดล (<https://umap-learn.readthedocs.io/en/latest/parameters.html>) ซึ่งเป็นเทคนิคในการลดมิติองค์ประกอบของข้อมูล (Dimension reduction technique) แบบหนึ่งเพื่อใช้ในการสร้างแบบจำลองหัวข้อ โดยกำหนดพารามิเตอร์ที่สำคัญคือ 1) วิธีการควบคุมความสมดุลของโครงสร้างของกลุ่มคำจากข้อมูล ($n_neighbors = 15$) โดยทั่วไปจะกำหนดไว้ที่ 15 2) การกำหนดระยะห่างน้อยที่สุดที่จะให้แต่ละคำให้อยู่ในมิติ/องค์ประกอบนั้น ($min_dist = 0.0$) สามารถกำหนดค่าพารามิเตอร์นี้ในช่วง 0.00-0.99 3) จำนวนหมวดหมู่/องค์ประกอบ ($n_components = 5$) และ 4) วิธีการคำนวณระยะห่างของข้อมูล ($metric = cosine$) โดยผลจากการวิเคราะห์แบบจำลองหัวข้อด้วยวิธี LDA จะแสดงจำนวนหัวข้อ (Topic) และคำสำคัญที่มีความถี่สูง 10 อันดับแรก พร้อมรายงานค่าโอกาสความเป็นได้ (Probability) ของคำนั้น ๆ ที่จะอยู่ในหัวข้อนั้น ๆ โดยหากมีค่าโอกาสความเป็นไปได้สูง จะแสดงถึงโอกาสที่คำนั้นจะอยู่ในหัวข้อนั้นสูงไปด้วย

4. การตั้งชื่อหมวดหมู่ของประเด็นหรือหัวข้อสำคัญ (Topic label) ผู้วิจัยได้พิจารณาถึงคำที่โดดเด่นของแต่ละประเด็น จากนั้นทำการจับกลุ่มว่าประเด็นสำคัญใดสอดคล้องกับเรื่องใดบ้างเพื่อนำไปสู่การตั้งชื่อหัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษา

5. การนำเสนอทัศนภาพข้อมูล (Data visualizations) ผู้วิจัยสร้างทัศนภาพข้อมูลเครือข่ายของคำด้วยโปรแกรม VOSviewer จากคลังข้อมูล (Corpus of documents) ชื่อเรื่องและบทคัดย่อ โดยที่ระยะห่างระหว่างโหนด (Node) สามารถบ่งชี้ความสัมพันธ์ของคำสำคัญที่ปรากฏจากผลการวิเคราะห์ ซึ่งหากระยะห่างเพิ่มมากขึ้นแปลความหมายได้ว่า ความสัมพันธ์ของคำสำคัญจะยิ่งลดน้อยลง โดยที่ระดับความสัมพันธ์ที่เกิดขึ้นนั้นเกิดจากผลการวิเคราะห์การปรากฏของคำร่วมกัน (Co-occurrences) (van Eck & Waltman, 2011) ตลอดจนการจำแนกกลุ่มของคำสำคัญสามารถพิจารณาได้จากสีที่แตกต่างกันของแต่ละโหนด และระดับความสำคัญของคำสำคัญสามารถพิจารณาได้จากขนาดของโหนดที่ปรากฏในทัศนภาพข้อมูลเครือข่ายของคำ

ผลการวิจัย (Results)

ผลการวิเคราะห์ข้อมูลทั่วไปของวิทยานิพนธ์ที่ใช้ในการศึกษา

ผลการวิเคราะห์ด้วยสถิติเชิงบรรยาย พบว่า จากข้อมูลวิทยานิพนธ์ทั้งหมดจำนวน 798 เรื่อง แบ่งเป็นภาษาไทย จำนวน 435 เรื่อง ภาษาอังกฤษ จำนวน 363 เรื่อง สามารถจำแนกตามปีที่มีการเผยแพร่ได้ดัง Table 1

Table 1

Dissertations Contribution by Years

ข้อมูลจำนวนวิทยานิพนธ์จำแนกตามปี

แหล่งวิทยานิพนธ์	ปี 2019 (%)	ปี 2020 (%)	ปี 2021 (%)	ปี 2022 (%)	ปี 2023 (%)
ประเทศไทย (435)					
จุฬาลงกรณ์มหาวิทยาลัย (435)	147 (33.793)	145 (33.333)	123 (28.275)	20 (4.597)	-
ประเทศสหรัฐอเมริกา (363)					
University of California, Los Angeles (203)	55 (27.094)	54 (26.601)	44 (21.675)	47 (23.153)	3 (1.477)
Harvard University (53)	2 (3.773)	4 (7.547)	19 (35.849)	28 (52.830)	-
Vanderbilt University (107)	29 (27.102)	37 (34.579)	22 (20.560)	19 (17.757)	-

ผลการวิเคราะห์เพื่อจัดหมวดหมู่วิทยานิพนธ์ทางการศึกษา

1. ผลการวิเคราะห์จากชุดข้อมูลชื่อเรื่อง พบว่า หัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาในประเทศไทย 5 หมวดหมู่แรก ประกอบด้วย 1) สุขภาวะและอารมณ์เชิงสังคม 2) กลยุทธ์การจัดการความคิดรวบยอด 3) กลยุทธ์การพัฒนาความสามารถทางคณิตศาสตร์ 4) การพัฒนาชุมชนการเรียนรู้ทางวิชาชีพ และ 5) การจัดการเรียนการสอนดนตรี ซึ่งหัวเรื่อง

แฝงที่มีความน่าจะเป็นสูงที่สุด ได้แก่ การจัดการเรียนการสอนดนตรี (Probability = .074) และหัวเรื่องแฝงของมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา 5 หมวดหมู่แรกประกอบด้วย 1) การพัฒนาผู้เรียนภาวะออทิซึม 2) ความแตกต่างทางอัตลักษณ์และชาติพันธุ์ 3) ความร่วมมือของชุมชน 4) วัฒนธรรมเชิงพื้นที่ และ 5) สุขภาวะผู้เรียนที่มีความผิดปกติทางร่างกาย ซึ่งหัวเรื่องแฝงที่มีความน่าจะเป็นสูงที่สุด ได้แก่ ความร่วมมือของชุมชน (Probability = .240) โดยมีรายละเอียดผลการวิเคราะห์ดัง Table 2

Table 2

List of 5 Latent Topics From Dissertation Titles Along With Their Keywords

หัวเรื่องแฝงของวิทยานิพนธ์ 5 หมวดหมู่แรก จากการวิเคราะห์ชื่อเรื่อง

Topic	ประเทศไทย			ประเทศสหรัฐอเมริกา		
	คำสำคัญ	การวิจัยด้าน	p	คำสำคัญ	การวิจัยด้าน	p
1	physical, student, effects, learning, program, social, emotional, fitness, school, education	สุขภาวะและอารมณ์เชิงสังคม	.021	learning, teacher, reading, mathematics, teachers, using, intervention, autism, study, english	การพัฒนาผู้เรียนภาวะออทิซึม	.051
2	concept, management, based, academic, strategy, school, innovation, secondary, eastern, corridor	กลยุทธ์การจัดการความคิดรวบยอด	.026	black, racial, school, identity, racism, experiences, students, engagement, race, college	ความแตกต่างทางอัตลักษณ์และชาติพันธุ์	.087

Table 2
(continued)

Topic	ประเทศไทย			ประเทศสหรัฐอเมริกา		
	คำสำคัญ	การวิจัยด้าน	p	คำสำคัญ	การวิจัยด้าน	p
3	mathematical, mathematics, ability, student, organizing, effects, using, grade, learning, scientific	กลยุทธ์การพัฒนาความสามารถทางคณิตศาสตร์	.017	college, community, education, essays, success, higher, career, faculty, colleges, financial	ความร่วมมือของชุมชน	.240
4	community, teacher, age, professional, development, learning, guidelines, lifelong, digital, model	การพัฒนาชุมชนการเรียนรู้ทางวิชาชีพ	.031	latina, angeles, los, school, education, high, latinx, higher, belonging, factors	วัฒนธรรมเชิงพื้นที่	.086
5	music, thai, song, teaching, piano, guidelines, guideline, violin, improvisation, choral	การจัดการเรียนการสอนดนตรี	.074	special, education, study, school, students, disabilities, health, inclusion, exploring, teachers	สุขภาวะผู้เรียนที่มี ความผิดปกติทางร่างกาย	.088

Figure 3

Latent Topics Analysis Data Visualization From Dissertation Titles

ทัศนภาพข้อมูลผลการวิเคราะห์หัวข้อแฝงจากชื่อเรื่องของวิทยานิพนธ์ทางการศึกษา

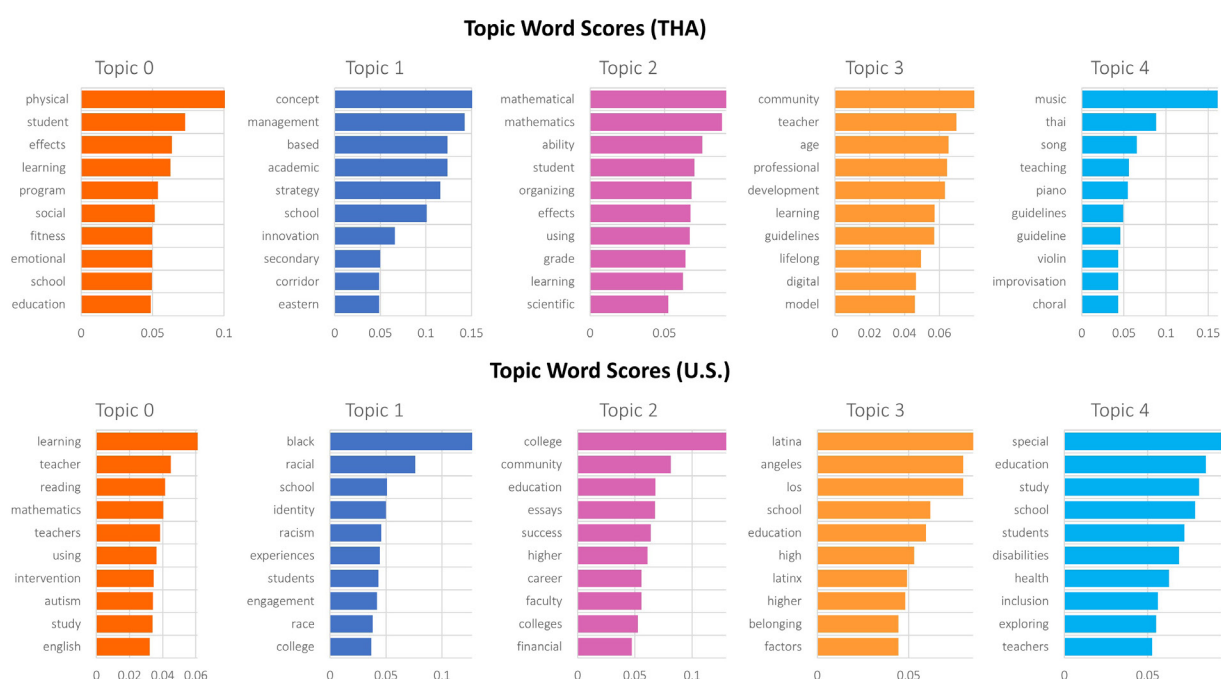


Figure 3
(continued)

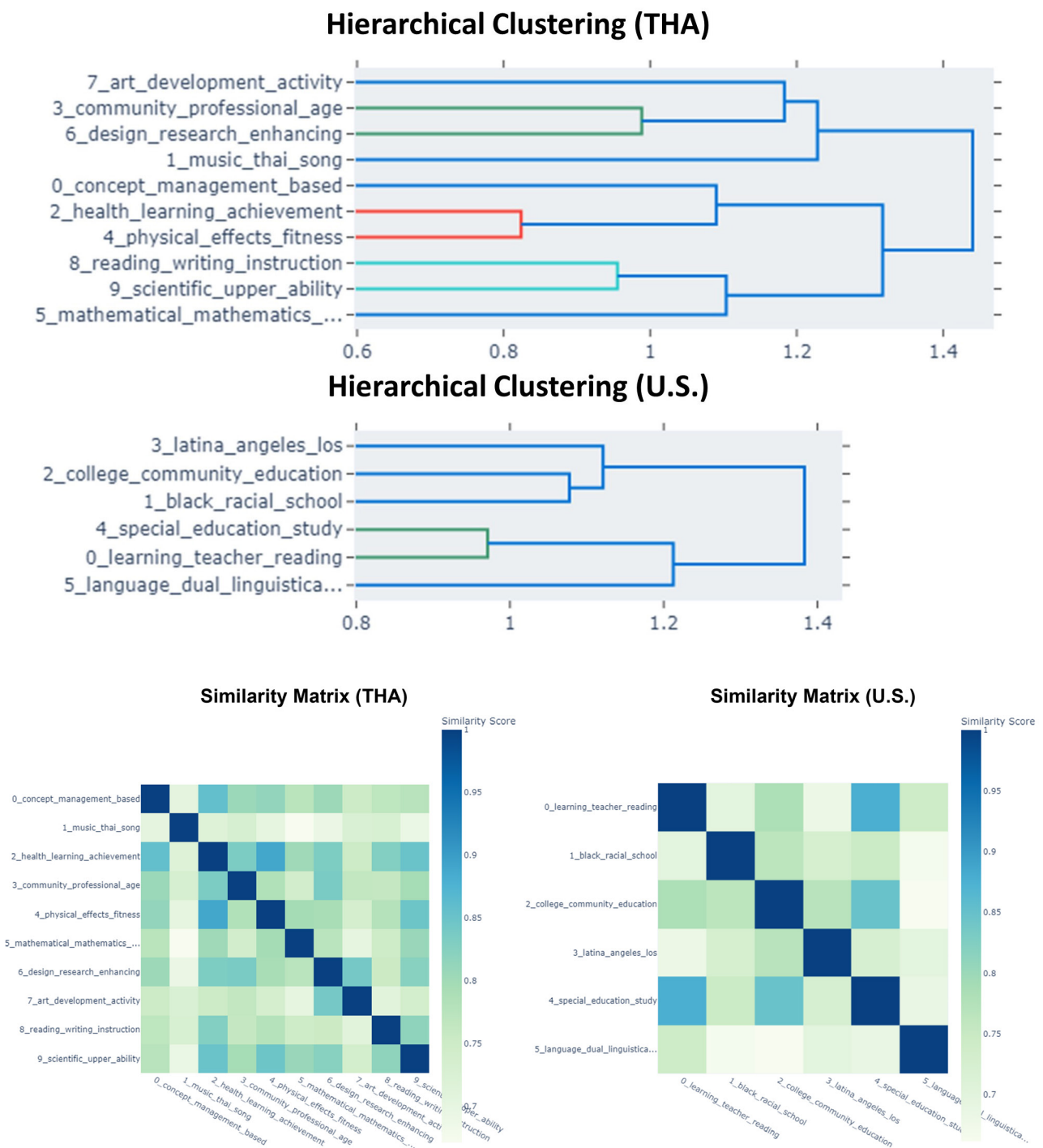


Figure 3 แสดงแผนภูมิแท่งคำสำคัญ 10 คำแรกในหัวเรื่องแต่ละหมวดหมู่ (Topic words scores) เรียงลำดับตามค่าคะแนน (Term scores) ซึ่งหากพิจารณาเปรียบเทียบค่าคะแนนระหว่างประเทศไทยและสหรัฐอเมริกาจะเห็นว่าไม่แตกต่างกันในแต่ละหัวเรื่องสำหรับผลการวิเคราะห์ความ

สัมพันธ์ของคำสำคัญระหว่างกลุ่มหัวเรื่องแฝง (Hierarchical clustering; Similarity matrix) แสดงให้เห็นว่าการกระจายตัวของคำสำคัญที่คล้ายคลึงกันระหว่างหัวเรื่องแฝงมากที่สุดได้แก่หัวเรื่องแฝงที่ 3 และ 5 สำหรับประเทศไทยและหัวเรื่องแฝงที่ 1 และ 5 สำหรับสหรัฐอเมริกา

Figure 4

VOSviewer Co-presence Structure of Words Across Latent Topics From Dissertation Titles

ทัศนภาพเครือข่ายของคำตามหัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาจากการวิเคราะห์เรื่องด้วย VOS viewer

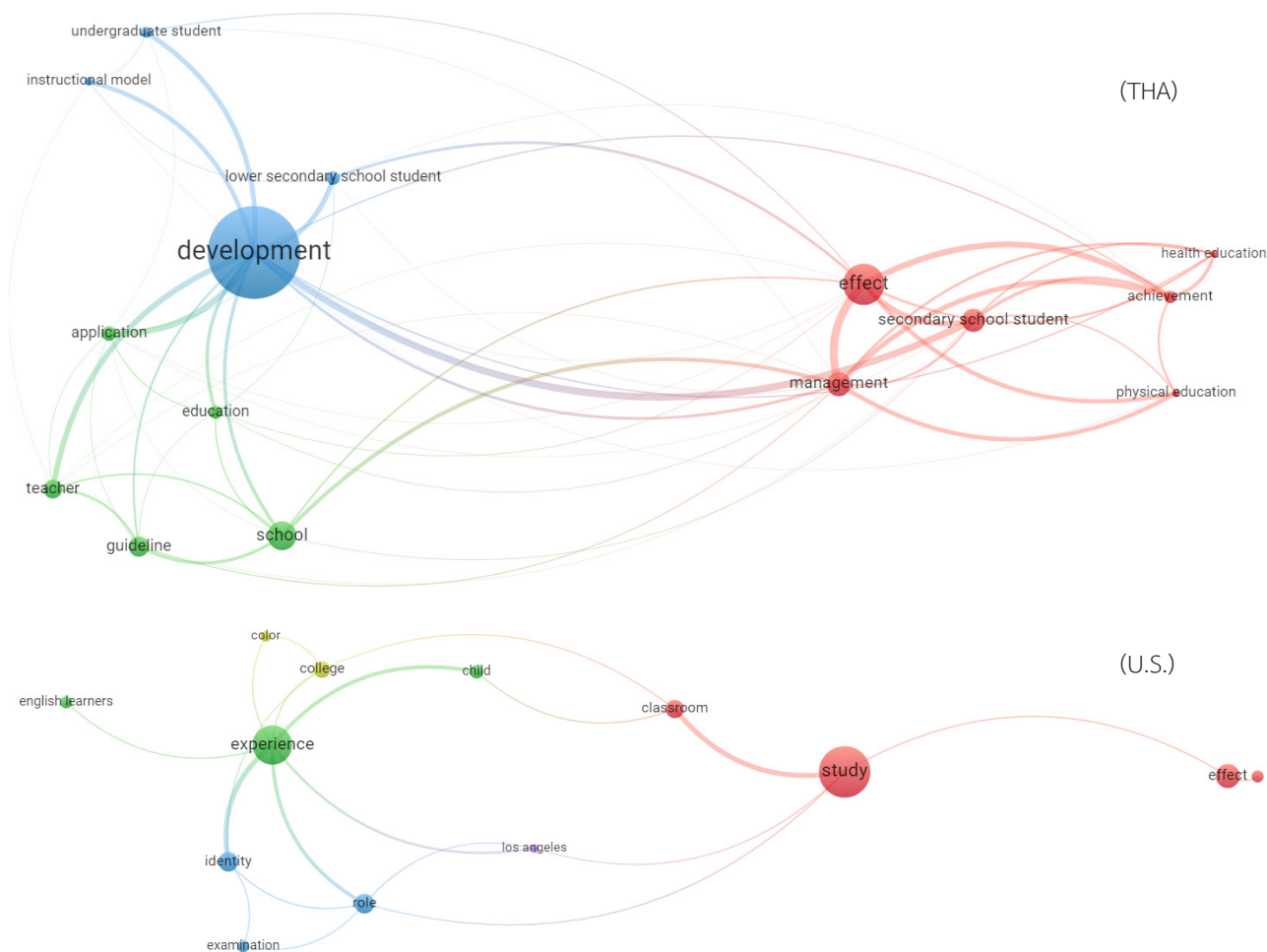


Figure 4 เมื่อพิจารณาจากผลการวิเคราะห์ของประเทศไทยและสหรัฐอเมริกาจะเห็นได้ว่าสามารถจัดกลุ่มคำสำคัญได้ 3 กลุ่มใหญ่เช่นเดียวกัน หากแต่ลักษณะโหนดของประเทศไทยมีขนาดใหญ่กว่าสหรัฐอเมริกา กล่าวคือ คำสำคัญ “development” มีความถี่ของการปรากฏของคำร่วมกันสูงในชื่อเรื่องของประเทศไทย สำหรับระดับความสัมพันธ์ของคำสำคัญที่ไม่แตกต่างกัน

2. ผลการวิเคราะห์จากบทคัดย่อ พบว่า หัวเรื่องแฝงในประเทศไทย 4 หมวดหมู่แรก ประกอบด้วย 1) โมเดลการเรียนรู้แบบกลุ่ม 2) กลยุทธ์การจัดการความคิดรวบยอด 3) การจัดการเรียนการสอนดนตรี และ 4) การจัดการเรียนการสอนศิลปะ

สร้างสรรค์ ซึ่งหัวเรื่องแฝงที่มีความน่าจะเป็นสูงที่สุด ได้แก่ การจัดการเรียนการสอนศิลปะสร้างสรรค์ (Probability = .798) สำหรับมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา 5 หมวดหมู่แรกประกอบด้วย 1) การพัฒนาผู้เรียนที่แตกต่างทางชาติพันธุ์ผ่านความร่วมมือของชุมชน 2) การพัฒนาครุคณิตศาสตร์ 3) อัตลักษณ์และชาติพันธุ์ของผู้เรียนเชิงพื้นที่ 4) การศึกษาสำหรับผู้เรียนสองภาษา และ 5) ความเท่าเทียมในสถานศึกษาของรัฐ ซึ่งหัวเรื่องแฝงที่มีความน่าจะเป็นสูงที่สุด ได้แก่ อัตลักษณ์และชาติพันธุ์ของผู้เรียนเชิงพื้นที่ (Probability = .343) โดยมีรายละเอียดผลการวิเคราะห์ดัง Table 3

Table 3

List of 5 Latent Topics From Dissertation Abstracts Along With Their Keywords

หัวข้อเรื่องแฝงของวิทยานิพนธ์ 5 หมวดหมู่แรก จากการวิเคราะห์บทคัดย่อ

Topic	ประเทศไทย			ประเทศสหรัฐอเมริกา		
	คำสำคัญ	การวิจัยด้าน	p	คำสำคัญ	การวิจัยด้าน	p
1	learning, student, group, research, using, level, score, teacher, model, study	โมเดลการเรียนรู้แบบกลุ่ม	.060	college, student, study, community, black, higher, students, research, faculty, education	การพัฒนาผู้เรียนที่แตกต่างทางชาติพันธุ์ผ่านความร่วมมือของชุมชน	.045
2	management, school, development, academic, evaluation, research, learning, concept, based, education	กลยุทธ์การจัดการความคิดรวบยอด	.037	teacher, learning, mathematics, group, teachers, practice, support, math, student, study	การพัฒนาครูคณิตศาสตร์	.163
3	music, song, violin, thai, research, content, performance, teaching, folk, book	การจัดการเรียนการสอนดนตรี	.044	latina, school, education, student, critical, study, community, studies, identity, high	อัตลักษณ์และชาติพันธุ์ของผู้เรียนเชิงพื้นที่	.343
4	art, activity, research, education, design, painting, product, cultural, creative, teaching	การจัดการเรียนการสอนศิลปะสร้างสรรค์	.798	language, english, home, study, classroom, student, spanish, vocabulary, writing, bilingual	การศึกษาสำหรับผู้เรียนสองภาษา	.047
5	-	-	-	district, school, learning, capstone, public, office, leader, system, equity, schools	ความเท่าเทียมในสถานศึกษาของรัฐ	.053

Figure 5

Latent Topics Analysis Data Visualization From Dissertation Abstracts

ทัศนภาพข้อมูลผลการวิเคราะห์หัวข้อแฝงจากบทคัดย่อของวิทยานิพนธ์ทางการศึกษา

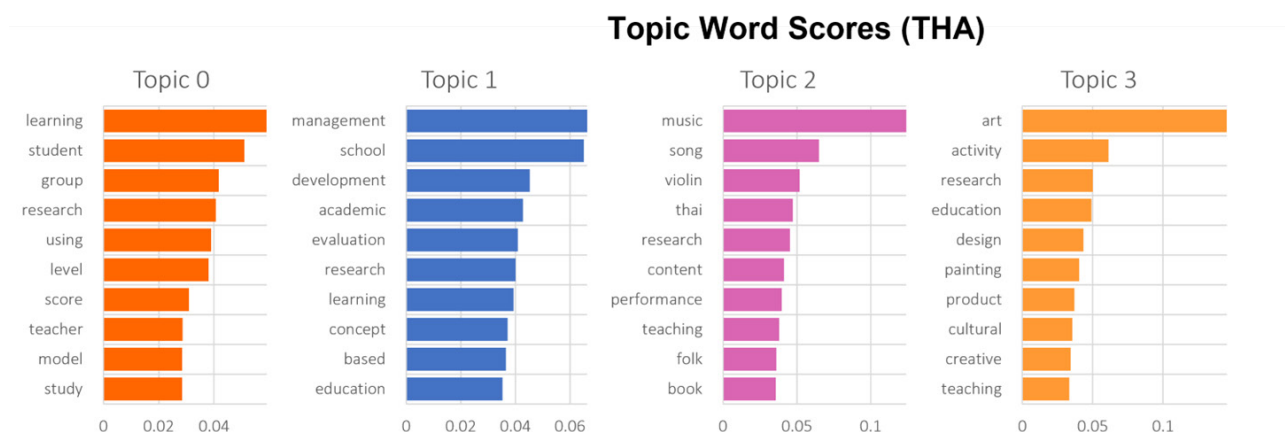
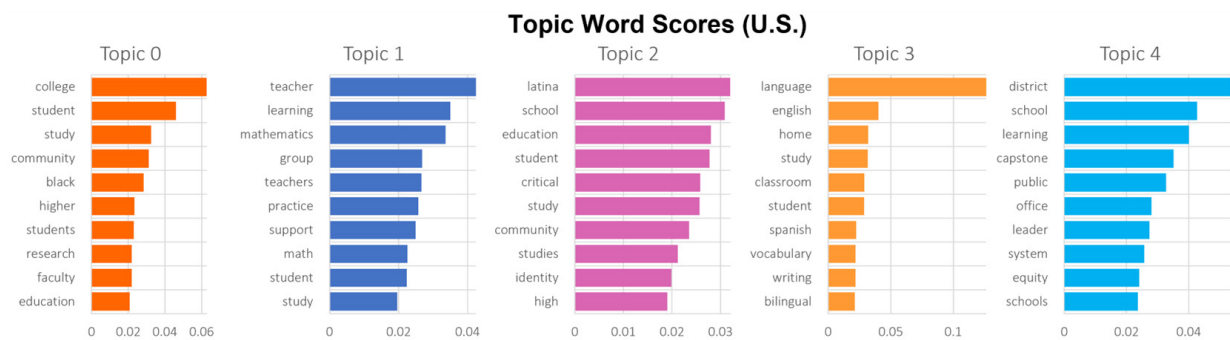
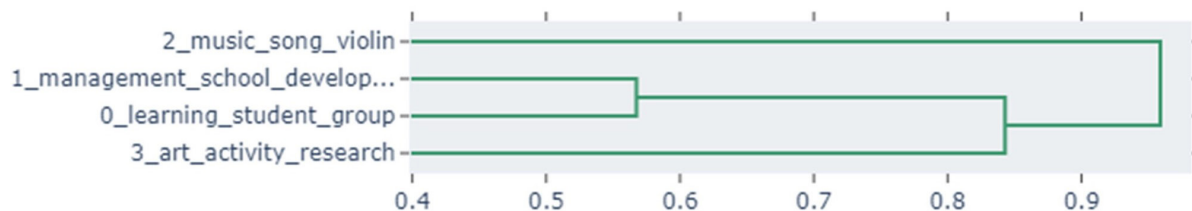


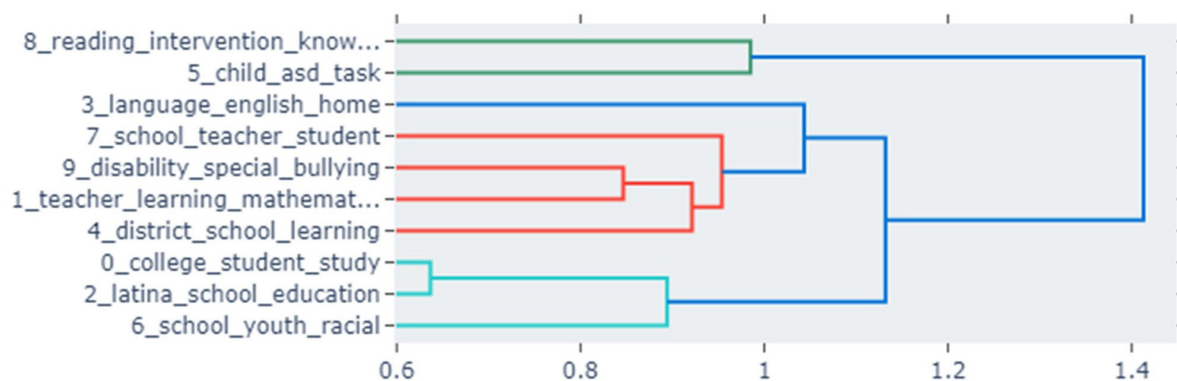
Figure 5
(continued)



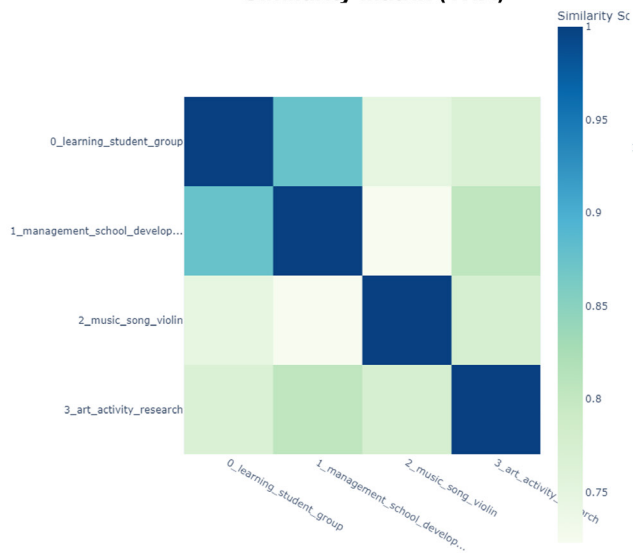
Hierarchical Clustering (THA)



Hierarchical Clustering (U.S.)



Similarity Matrix (THA)



Similarity Matrix (U.S.)

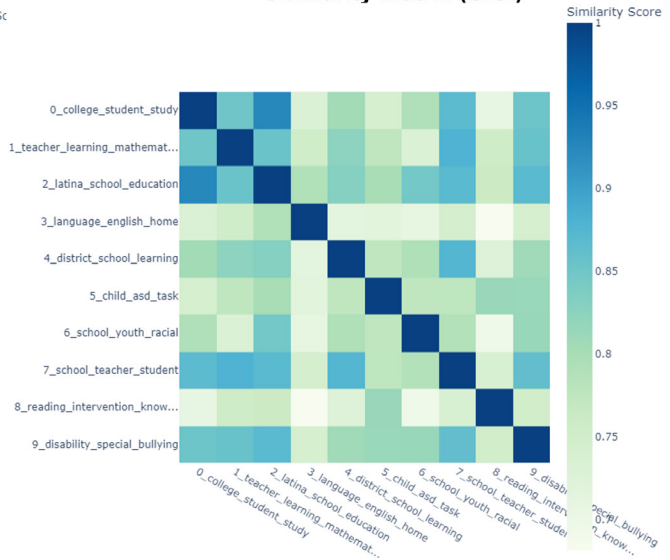


Figure 5 แสดงแผนภูมิแท่งค่าสำคัญ 10 คำแรก ในหัวเรื่องแฝงแต่ละหมวดหมู่เรียงลำดับตามค่าคะแนนที่ได้ จากการวิเคราะห์จากบทคัดย่อ ซึ่งหากพิจารณาเปรียบเทียบ ค่าคะแนนระหว่างประเทศไทยและสหรัฐอเมริกาจะเห็นว่า ไม่แตกต่างกันในแต่ละหัวเรื่องแฝง สำหรับผลการวิเคราะห์

ความสัมพันธ์แสดงให้เห็นว่าการกระจายตัวของค่าสำคัญที่คล้ายคลึงกันระหว่างหัวเรื่องแฝงมากที่สุดได้แก่ หัวเรื่องแฝงที่ 1 และ 2 สำหรับประเทศไทย และหัวเรื่องแฝงที่ 1 และ 3 สำหรับสหรัฐอเมริกา

Figure 6

VOSviewer Co-presence Structure of Words Across Latent Topics From Dissertation Abstract

ทัศนภาพเครือข่ายของคำตามหัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาจากการวิเคราะห์ที่เชื่อมทศด้วย VOS viewer

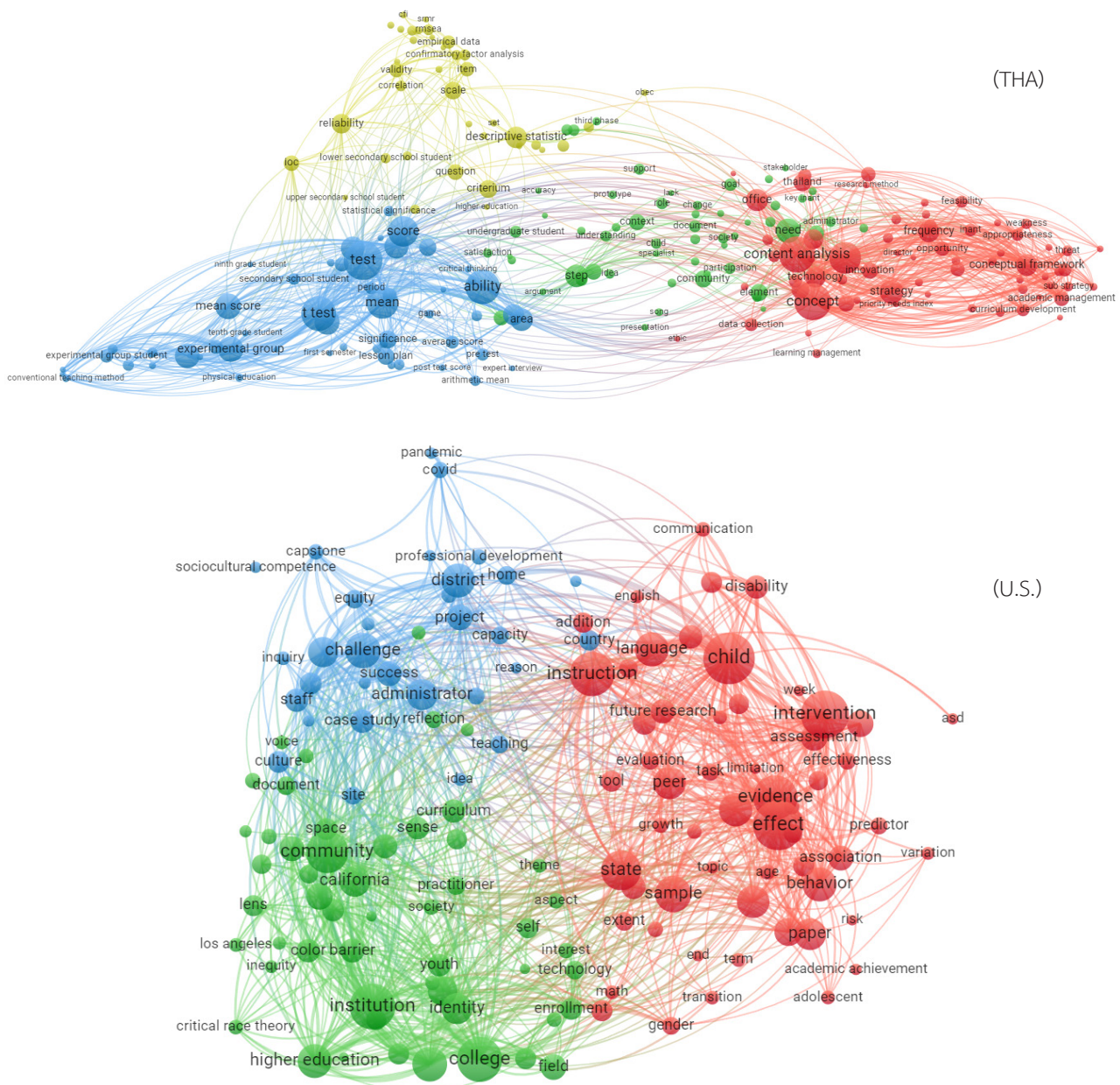


Figure 6 เมื่อพิจารณาจากผลการวิเคราะห์ของประเทศไทยสามารถจัดกลุ่มคำสำคัญได้ 4 กลุ่มใหญ่ ซึ่งแตกต่างจากสหรัฐอเมริกาที่สามารถจัดกลุ่มคำสำคัญได้ 3 กลุ่มใหญ่ตลอดจนความแตกต่างกันในประเด็นระหว่างประเทศ

แสดงให้เห็นว่า สหรัฐอเมริกามีระดับความสัมพันธ์ระหว่างคำ
สำคัญสูงกว่าประเทศไทยซึ่งพิจารณาจากระยะห่างระหว่าง
โหนดที่ปรากฏ

ผลการวิเคราะห์เพื่อเปรียบเทียบลักษณะหัวเรื่องแฝงของหัวข้อวิทยานิพนธ์ทางการศึกษาจากการวิเคราะห์ด้วยชื่อเรื่องและบทคัดย่อ

จากผลการวิเคราะห์เพื่อจัดหมวดหมู่วิทยานิพนธ์ทางการศึกษาด้วยวิธีการทำ Topic modeling สามารถจำแนกประเด็นสำหรับการเปรียบเทียบลักษณะหัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาได้ดังนี้

1. สำหรับประเทศไทย พบว่า หัวเรื่องแฝงที่ได้จากการวิเคราะห์ชื่อเรื่องมีการจัดเป็นหมวดหมู่ที่มีความละเอียดมากกว่าการวิเคราะห์จากบทคัดย่อ โดยมีความคล้ายคลึงและแตกต่างกันดังนี้ หมวดหมู่ที่มีความคล้ายกัน ได้แก่ หัวเรื่องแฝงที่ดำเนินการวิจัยด้านการจัดการเรียนการสอนดนตรี ซึ่งเป็นหัวเรื่องแฝงที่มีความน่าจะเป็นสูงที่สุดจากการวิเคราะห์ด้วยบทคัดย่อและชื่อเรื่องตามลำดับ รวมถึงในการวิจัยด้านกลยุทธ์การจัดการความคิดรวบยอดสำหรับการศึกษา ในขณะที่มีความแตกต่างกันในหลายหมวดหมู่ เช่น การวิจัยด้านการพัฒนาชุมชนการเรียนรู้ทางวิชาชีพ ซึ่งถูกแยกออกมาเป็นหนึ่งหมวดหมู่เมื่อทำการวิเคราะห์จากชื่อเรื่อง หรือการวิจัยด้านสุขภาวะและอารมณ์เชิงสังคมที่พบจากการวิเคราะห์จากชื่อเรื่อง แต่ไม่ปรากฏในการวิเคราะห์จากบทคัดย่อ

2. สำหรับมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา พบว่า หัวเรื่องแฝงที่ได้จากการวิเคราะห์ชื่อเรื่องมีการจัดเป็นหมวดหมู่ใหญ่ ๆ ในขณะที่ผลการวิเคราะห์จากบทคัดย่อจะแสดงหมวดหมู่ย่อย ๆ ที่มีความละเอียดมากกว่า นอกจากนี้ ยังพบว่า มีความเหมือนและแตกต่างกันในบางหมวดหมู่ โดยหมวดหมู่ที่มีความคล้ายคลึงกัน เช่น วิทยานิพนธ์ที่ดำเนินการวิจัยด้านอัตลักษณ์ เชื้อชาติ ภาษาศาสตร์ การพัฒนาผู้เรียนที่มีความแตกต่างกันอย่างเท่าเทียม และมีความแตกต่างกันในบางหมวดหมู่ อย่างไรก็ตาม หมวดหมู่ที่มีความน่าจะเป็นสูงที่สุดจากการวิเคราะห์จากชื่อเรื่องและบทคัดย่อมีความสอดคล้องไปในทิศทางเดียวกัน คือ การวิจัยที่เกี่ยวข้องกับการให้ความสำคัญสำหรับการจัดการศึกษาบนพื้นฐานของความหลากหลายของผู้เรียน

3. เมื่อเปรียบเทียบการวิเคราะห์เพื่อจัดหมวดหมู่วิทยานิพนธ์ พบว่า หัวเรื่องแฝงมีความแตกต่างกันเป็นส่วนใหญ่ โดยการทำวิทยานิพนธ์ในประเทศไทยมุ่งเน้นไปที่การวิจัยด้านการศึกษาวิธีการหรือกลยุทธ์การจัดการเรียนการสอนเพื่อพัฒนาคุณภาพผู้เรียน แตกต่างจากมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาที่มีการวิจัยในประเด็นที่หลากหลายนอกเหนือจากการศึกษาวิธีการจัดการเรียนการสอน โดยมุ่งเน้นไปที่การวิจัยในประเด็นที่เกี่ยวข้องกับบทบาทของชุมชนสำหรับการส่งเสริมการจัดการศึกษา นอกจากนี้ยังพบการวิจัยประเด็นการจัดการเรียนการสอนที่ให้ความสำคัญกับความแตกต่างของผู้เรียน ซึ่งไม่ปรากฏในผลการวิเคราะห์วิทยานิพนธ์ทางการศึกษาในประเทศไทย

อภิปรายผล (Discussions)

การศึกษาลักษณะหัวเรื่องแฝงของวิทยานิพนธ์ทางการศึกษาในประเทศไทยและมหาวิทยาลัยชั้นนำในสหรัฐอเมริกา พบประเด็นที่น่าสนใจในการอภิปรายดังต่อไปนี้

การใช้แนวคิดการสร้างแบบจำลองหัวข้อเพื่อทำการวิเคราะห์หัวข้อแฝงเป็นรูปแบบหนึ่งของการวิเคราะห์เหมืองข้อความ ซึ่งถือเป็นวิธีการวิเคราะห์ข้อมูลแบบใหม่ผ่านการใช้เครื่อง (Machine) ในการวิเคราะห์ข้อมูล ซึ่งมีความแตกต่างจากการวิเคราะห์ข้อความแบบดั้งเดิมหรือการวิเคราะห์เนื้อหาในประเด็นสำคัญต่อไปนี้ วิธีการวิเคราะห์แบบใหม่เป็นแนวคิดวิธีการวิจัยเชิงปริมาณ ข้อมูลที่ใช้ในการวิเคราะห์จะอยู่ในรูปของข้อความ หน่วยในการวิเคราะห์ ได้แก่ หน่วยคำ (Words) หน่วยคำย่อย (Stems) และการรวมกันของคำ (Combination of words) ผ่านการวิเคราะห์โดยเครื่องซึ่งเป็นการวิเคราะห์โดยไม่มีกรอบทฤษฎีตั้งต้น (Theory-free) ผลการวิเคราะห์มักจะถูกนำเสนอในลักษณะของตารางและกราฟ ในขณะที่วิธีการวิเคราะห์แบบดั้งเดิมจะเป็นวิธีการวิจัยเชิงคุณภาพเป็นหลัก ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลเชิงคุณภาพ เช่น ข้อความ รูปภาพ วิดีโอ มีหน่วยในการวิเคราะห์ เช่น คำ ประโยค ทำการวิเคราะห์ข้อมูลโดยนักวิจัยเป็นหลัก ซึ่งในการทำงานที่มีข้อมูลจำนวนมากจะทำให้เกิดการใช้ทรัพยากรที่มากตามไปด้วย เช่น แรงงาน ค่าใช้จ่าย รวมทั้งยังอาจเกิดความผิดพลาดขึ้นในกระบวนการวิเคราะห์ข้อมูลได้ การวิเคราะห์แบบดั้งเดิมมักจะมีกรอบทฤษฎีกำหนดไว้ ผลการวิเคราะห์ส่วนใหญ่จะถูกนำเสนอในลักษณะของข้อมูลที่มีอ้างอิง ตาราง และกราฟ (Aureli, 2017)

การศึกษาค้นคว้านี้มีวัตถุประสงค์หลักเพื่อวิเคราะห์หัวเรื่องแฝงและจัดหมวดหมู่วิทยานิพนธ์ทางการศึกษา ซึ่งเป็นการใช้วิธีการวิเคราะห์แบบใหม่ที่สามารถใช้กับข้อมูลที่มีจำนวนมากได้อย่างมีประสิทธิภาพ และจากการศึกษางานวิจัยที่เกี่ยวข้องในประเทศ พบว่า มีการใช้วิธีการวิเคราะห์แบบใหม่ในการวิเคราะห์ข้อมูลในลักษณะที่ใกล้เคียงกัน โดย Noklang (2019) ได้ทำการศึกษาเพื่อวิเคราะห์ความสอดคล้องระหว่างคุณลักษณะของหลักสูตร และคุณลักษณะของวิทยานิพนธ์สาขาวิจัยการศึกษาของไทยและต่างประเทศ รวมทั้งเปรียบเทียบความสอดคล้องระหว่างคุณลักษณะของหลักสูตรและคุณลักษณะของวิทยานิพนธ์สาขาวิจัยการศึกษาของไทยและต่างประเทศ โดยใช้ข้อมูลหลักสูตรและวิทยานิพนธ์ที่รวบรวมได้จากฐานข้อมูลออนไลน์ และใช้การวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อความ วิเคราะห์เครือข่ายข้อความ (Text network) และการวิเคราะห์สมนัย (Correspondence analyses)

นอกจากนั้น ยังพบงานวิจัยในต่างประเทศที่ใช้การวิเคราะห์ด้วยวิธีการใหม่เพื่อศึกษาและเปรียบเทียบข้อมูลด้านการศึกษา เช่น การศึกษาของ Murrugarra-Llerena et al.

(2022) ที่ใช้การวิเคราะห์ของเครื่องเพื่อเปรียบเทียบหลักสูตรคอมพิวเตอร์ศึกษาในระดับอุดมศึกษา โดยใช้ข้อมูลในการวิเคราะห์ ได้แก่ หลักสูตรคอมพิวเตอร์ศึกษาของสหรัฐอเมริกาจำนวน 296 หลักสูตรและกลุ่มประเทศละตินจำนวน 18 หลักสูตรทำการวิเคราะห์และแสดงผลในลักษณะของ Word clouds เพื่อเปรียบเทียบความถี่ของคำที่พบจากแต่ละหลักสูตร รวมถึงใช้การวิเคราะห์ด้วยโมเดล BERT เพื่อจัดหมวดหมู่ของหลักสูตร ตลอดจนนักวิชาการที่ศึกษาการวิเคราะห์ของเครื่องด้วยโมเดลการวิเคราะห์ด้วย LDA ดังเช่น Karbasian and Johri (2020) ศึกษาการสร้างแบบจำลองหัวข้อของข้อความจากการแสดงความคิดเห็นด้านวิทยาการข้อมูล (Data science) บนแพลตฟอร์มออนไลน์ชื่อ “Stack Exchange” และ “Reddit” ทำการเปรียบเทียบหมวดหมู่ของการแสดงความคิดเห็นในแต่ละแพลตฟอร์ม ผลการวิจัยที่ได้ยังนำไปสู่การพัฒนาและปรับปรุงหลักสูตรและการจัดการเรียนการสอนในส่วนที่เกี่ยวข้องกับวิทยาการข้อมูล ในขณะที่ Shrader et al. (2021) สร้างแบบจำลองหัวข้อแฝงจากข้อมูลการสัมภาษณ์นักศึกษาระดับบัณฑิตศึกษาสาขาอาชีพและการตรวจสอบบัญชี จำนวน 26 คน ในประเด็นที่เกี่ยวข้องกับการโกงในสาขาวิชาชีพ ผลการวิเคราะห์แสดงให้เห็นหมวดหมู่ของหัวข้อที่เกี่ยวข้องกับการโกงในประเด็นต่าง ๆ เช่น ประเด็นทั่วไปในการทำงาน (General context and code) การมอบหมายงาน (Assignment) การทำงานร่วมกัน (Teams)

จากผลการเปรียบเทียบการวิเคราะห์ พบว่า หมวดหมู่วิทยานิพนธ์มีความแตกต่างกันเป็นส่วนใหญ่ เนื่องจากในประเทศไทยส่วนใหญ่เป็นการวิจัยที่มุ่งเน้นการพัฒนาการจัดการเรียนการสอนการส่งเสริมและพัฒนาการเรียนรู้ของผู้เรียน รวมถึงพัฒนาครูผู้สอนเป็นสำคัญ เนื่องด้วยประเทศไทยให้ความสำคัญกับการพัฒนาคุณภาพของครู ผู้เรียน และกระบวนการจัดการเรียนการสอน ซึ่งเชื่อว่าเป็นส่วนสำคัญที่จะทำให้การศึกษาเกิดการพัฒนาดังอย่างมีคุณภาพและยั่งยืน ซึ่งมีความสอดคล้องกับงานวิจัยภาพการศึกษาไทยในอนาคต 10-20 ปี ของ Siridhrungsri (2009) ที่พบว่า การจัดการศึกษาของประเทศไทยจะหันมามุ่งเน้นในเรื่องของคุณภาพมากขึ้น ทั้งคุณภาพครู คุณภาพผู้เรียน และคุณภาพของระบบการจัดการศึกษา ซึ่งหมายถึงประสิทธิภาพและประสิทธิผลในการบริหารจัดการ เพื่อให้คนไทยเป็นคนคุณภาพและศูนย์กลางการพัฒนาเพื่อนำไปสู่การพัฒนาประเทศได้อย่างแท้จริง จึงเห็นได้ว่าทั้งงานวิจัยในปัจจุบันและอนาคตอันใกล้ของประเทศไทยมีแนวโน้มที่จะเป็นการทำวิจัยเพื่อมุ่งเน้นหาวิธีการหรือแนวทางในการพัฒนาครู ผู้เรียน และคุณภาพของการจัดการศึกษาเป็นสำคัญ

ในขณะมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาที่มีประเด็นในการศึกษาที่หลากหลาย มุ่งเน้นไปที่การศึกษาในประเด็นที่เกี่ยวข้องกับการศึกษาในเชิงพื้นที่ รวมถึงการศึกษาในประเด็น

ที่เกี่ยวกับอัตลักษณ์และชาติพันธุ์ ตลอดจนความเท่าเทียมทางการศึกษา ซึ่งอาจเป็นผลมาจากในประเทศสหรัฐอเมริกาที่มีประชากรจากหลากหลายชาติพันธุ์อาศัยอยู่ ซึ่งย่อมส่งผลให้เกิดความแตกต่างในแง่ของโอกาสในการเข้าถึงการศึกษาและโอกาสอื่น ๆ ในสังคม นักวิชาการบางส่วนเชื่อว่าระบบการศึกษาที่เท่าเทียมกันจะเข้ามาช่วยแก้ปัญหาความเหลื่อมล้ำที่เกิดขึ้นได้ระหว่างกลุ่มชาติพันธุ์ได้ โดย Hanauer (2019) ได้กล่าวถึง การที่ความยากจนและความเหลื่อมล้ำเพิ่มสูงขึ้นในสหรัฐอเมริกาเป็นผลมาจากระบบการศึกษาที่ล้มเหลว และได้เสนอให้แก้ปัญหาดังกล่าวด้วยการพัฒนาระบบการศึกษาในอเมริกาให้มีคุณภาพและเท่าเทียมมากขึ้น นอกจากนี้ Hamilton (2020) ได้ให้ความเห็นว่า ความเหลื่อมล้ำทางการศึกษาเป็นเรื่องเกี่ยวกับความเหลื่อมล้ำในการเข้าถึงทรัพยากรทางการศึกษาระหว่างกลุ่มสังคมต่าง ๆ ซึ่งมักเป็นผลมาจากนโยบายของรัฐบาล การเลือกโรงเรียน ความมั่งคั่งของครอบครัวที่ตั้งอยู่อาศัย แนวทางในการเลี้ยงดู และทางเลือก ตลอดจนอคติโดยนัยต่อเชื้อชาติ ชาติพันธุ์ และเพศของนักเรียน ดังนั้น หมวดหมู่วิทยานิพนธ์ทางการศึกษาของมหาวิทยาลัยชั้นนำในสหรัฐอเมริกาจึงเป็นการศึกษาในประเด็นที่เกี่ยวข้องกับการพัฒนานโยบาย หรือการจัดการเรียนการสอนที่เท่าเทียมกันในทุกกลุ่มชาติพันธุ์ รวมถึงกลุ่มที่มีความแตกต่างทางด้านเศรษฐกิจและสังคม และยังหมายรวมถึงผู้เรียนที่มีความต้องการพิเศษในด้านต่าง ๆ ซึ่งพบว่า มีนักเรียนที่มีความต้องการพิเศษถึง 7.3 ล้านคนในสหรัฐอเมริกา หรือคิดเป็นร้อยละ 15 ของการลงทะเบียนเรียนในโรงเรียนรัฐระหว่างปีการศึกษา 2021-2022 (Schaeffer, 2023) ซึ่งสะท้อนให้เห็นถึงความท้าทายในการจัดการศึกษาให้กับผู้เรียนที่มีความต้องการพิเศษเพื่อตอบสนองความต้องการที่หลากหลาย โดยต้องคำนึงถึงสุขภาพกายและสุขภาพจิตของผู้เรียนกลุ่มนี้เป็นสำคัญด้วย จึงเป็นเรื่องยากในการจัดการศึกษาให้สอดคล้องและเหมาะสมกับผู้เรียนแต่ละบุคคล จึงน่าจะเป็นอีกเหตุผลหนึ่งที่ทำให้แนวโน้มในการทำวิจัยเป็นการศึกษาเพื่อหาวิธีการและแนวทางในการจัดการศึกษาที่เหมาะสมที่สุดสำหรับผู้เรียนกลุ่มนี้

■ ข้อเสนอแนะ (Recommendations)

การศึกษาค้นคว้าครั้งนี้มุ่งศึกษาเกี่ยวกับแบบจำลองหัวข้อจากวิทยานิพนธ์ของมหาวิทยาลัยต่างประเทศและในประเทศไทยที่มุ่งกำหนดคำค้นโดยกว้างที่แสดงประเด็นการวิจัยที่มีการศึกษาในภาพกว้างสำหรับวิทยานิพนธ์ทางการศึกษา ตลอดจนช่องว่างในการศึกษาวิจัยเกี่ยวกับการศึกษา และลักษณะวิทยานิพนธ์เฉพาะตามบริบทของสภาพสังคมและชุมชน สำหรับในการวิจัยครั้งต่อไปมีข้อเสนอแนะ ดังนี้

1. ในการทำวิจัยหรือวิทยานิพนธ์ทางการศึกษาผู้ที่สนใจอาจเริ่มต้นด้วยการค้นคว้าประเด็นที่ต้องการศึกษาในขอบเขต

เฉพาะ เช่น การใช้เทคโนโลยีสำหรับการศึกษา การศึกษาในยุคที่มีการแพร่ระบาดใหญ่ (COVID-19) เพื่อศึกษาทิศทางและแนวโน้มการเปลี่ยนแปลงของหัวข้องานวิจัยที่จะศึกษาและมีความทันสมัย สอดคล้องกับการเปลี่ยนแปลง และตอบสนองต่อความต้องการของผู้เรียนในยุคปัจจุบัน

2. การนำวิธีวิทยาการวิจัยทางเหมืองข้อมูล (Data mining) มาประยุกต์ใช้สำหรับการศึกษาวิจัยสำหรับข้อมูลทางการศึกษาที่มีจำนวนมากเพื่อศึกษาค้นคว้าทิศทางแนวโน้มการเปลี่ยนแปลงและช่องว่างสำหรับศึกษาวิจัยของหัวข้องานวิจัยใน Scope fields ที่สนใจศึกษา

เอกสารอ้างอิง (References)

- Al-Rawi, A., Al-Musalli, A., & Fakida, A. (2021). News values on Instagram: A comparative study of international news. *Journalism and Media*, 2(2), 305-320. <https://doi.org/10.3390/journalmedia2020018>
- Amara, A., Hadj Taieb, M. A., & Ben Aouicha, M. (2021). Multilingual topic modeling for tracking COVID-19 trends based on Facebook data analysis. *Applied Intelligence*, 51, 3052-3073. <https://doi.org/10.1007/s10489-020-02033-3>
- Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *International Journal of Digital Accounting Research*, 17, 1-32. http://doi.org/10.4192/1577-8517-v17_1
- Bhattacharya, D. Sahoo, S., & Panda, B. N. (2021). *Emerging educational research trends in 21st century*. ResearchGate. https://www.researchgate.net/publication/359930009_Emerging_Educational_Research_Trends_in_21st_Century
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(4/5), 993-1022. <https://search.ebscohost.com/login.aspx?direct=true&db=asn&AN=12323372&site=ehost-live&scope=site>
- Chauhan, U., & Shah, A. (2021). Topic modeling using latent dirichlet allocation: A survey. *ACM Computing Surveys*, 54(7), 1-35. <https://doi.org/10.1145/3462478>
- Ekin, C. C., Polat, E., & Hopcan, S. (2023). Drawing the big picture of games in education: A topic modeling-based review of past 55 years. *Computers & Education*, 194, 104700. <https://doi.org/10.1016/j.compedu.2022.104700>
- Hamilton, R. N. (2020, December 15). *Educational inequality in America: race and gender*. Around Robin. <https://www.aroundrobin.com/educational-inequality-in-america>
- Hanauer, N. (2019, July 15). *Better schools won't fix America*. The Atlantic. <https://www.theatlantic.com/magazine/archive/2019/07/education-isnt-enough/590611>
- Hong, L., & Davison, B. D. (2010, July 25-28). *Empirical study of topic modeling in twitter* [Conference session]. The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D.C., USA. <https://doi.org/10.1145/1964858.1964870>
- Hwang, S., Flavin, E., & Lee, J. E. (2023). Exploring research trends of technology use in mathematics education: A scoping review using topic modeling. *Education and Information Technologies*, 1-28. <https://doi.org/10.1007/s10639-023-11603-0>
- Kaemane, T. (2010). *Sat kanson: 'ongkhwamru phra kanchat krabuan kan ranru thi mi prasitthipap* [Science of teaching: Knowledge for effective learning organization]. Chulalongkorn University.
- Karbasian, H., & Johri, A. (2020, March 11-14). *Insights for curriculum development: Identifying emerging data science topics through analysis of Q&A communities* [Conference session]. SIGCSE '20: The 51st ACM Technical Symposium on Computer Science Education, Oregon, USA. <https://doi.org/10.1145/3328778.3366817>
- Mazumder, S., & Barui, T. (2021). Discovering topics from the titles of the Indian LIS theses. *Library Philosophy and Practice (e-journal)*, 1, 5924. <https://digitalcommons.unl.edu/libphilprac/5924>
- Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, 11(18), 8438. <https://doi.org/10.3390/app11188438>
- Murrugarra-Llerena, J., Alva-Manchego, F., & Murrugarra-Llerena, N. (2022, December 7-11). *Improving embeddings representations for comparing higher education curricula: A use case in computing* [Conference session]. The 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates. <http://doi.org/10.18653/v1/2022.emnlp-main.776>
- Noklang, S. (2019). *Alignment of curricula and theses in educational research graduate programs: Application of text mining of Thai and international databases* [Doctoral dissertation, Chulalongkorn University]. CUIR Database. <https://cui.car.chula.ac.th/handle/123456789/70021>
- The QS World University Rankings. (2021). *QS World university rankings by subject 2021: Education & Training*. The QS World University Rankings. <https://www.topuniversities.com/university-rankings/university-subject-rankings/2021/education-training?®ion=Asia&countries=th>
- Schaeffer, K. (2023, July 24). *What federal education data shows about students with disabilities in the U.S.* Pew Research Center. <https://www.pewresearch.org/short-reads/2023/07/24/what-federal-education-data-shows-about-students-with-disabilities-in-the-us>
- Schmiedel, T., Müller, O., & Vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, 22(4), 941-968. <https://doi.org/10.1177/1094428118773858>
- Sheridan, S. (2022, November 16). *What is topic modeling? A beginner's guide*. Levity. <https://levity.ai/blog/what-is-topic-modeling>
- Shrader, C. B., Ravenscroft, S. P., Kaufmann, J. B., & Hansen, K. (2021). Collusion among accounting students: Data visualization and topic modeling of student interviews. *Decision Sciences Journal of Innovative Education*, 19(1), 40-62. <https://doi.org/10.1111/dsji.12226>
- Siridhrungsri, P. (2009). *Phap kansuksa Thai nai 'anakhot sip-yisip pi: Raingān kanwichai* [Thailand education scenario in 10-20 years: research report]. Office of the Educational Council.
- The Times Higher Education World University Rankings. (2020). *World university rankings 2020 by subject*. The Times Higher Education World University Rankings. https://www.timeshighereducation.com/world-university-rankings/2020/world-ranking#/length/25/locations/THA/subjects/3108/sort_by/rank/sort_order/asc/cols/stats
- U.S. News & World Report. (2023). *2023-2024 Best education schools*. U.S. News & World Report. <https://www.usnews.com/best-graduate-schools/top-education-schools/edu-rankings>
- van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. arXiv preprint arXiv: 1109.2058. <https://doi.org/10.48550/arXiv.1109.2058>
- Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, 94, 1-15. <https://doi.org/10.1016/j.is.2020.101582>
- Walker, R. M., Chandra, Y., Zhang, J., & Van Witteloostuijn, A. (2019). Topic modeling the research practice gap in public administration. *Public Administration Review*, 79(6), 931-937. <https://doi.org/10.1111/puar.13095>
- Yang, T. I., Torget, A., & Mihalcea, R. (2011, June 24). *Topic modeling on historical newspapers* [Conference session]. The 5th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Oregon, USA. <https://aclanthology.org/W11-1513.pdf>
- Yun, E. (2020). Review of trends in physics education research using topic modeling. *Journal of Baltic Science Education*, 19(3), 388-400. <http://dx.doi.org/10.33225/jbse/20.19.388>