

## บทที่ 6

### สรุปผลการวิจัย

ในบทนี้จะแสดงการสรุปผลการทดลองของงานวิจัย ปัญหาที่เกิดขึ้นในการทดลอง และข้อเสนอแนะของงานวิจัย เพื่อปรับเปลี่ยนรูปแบบของงานวิจัยหรือพัฒนาการทดลองให้มีประสิทธิภาพยิ่งขึ้น

#### 6.1 การทดลองและลักษณะของข้อมูลที่ใช้ทดสอบการค้นคืนเอกสาร

งานวิจัยนี้เป็นงานวิจัยเชิงประจักษ์ (Empirical Research) ซึ่งเป็นการศึกษาทดสอบระบบการค้นคืนเอกสารที่ประยุกต์ใช้ทฤษฎี เพื่ออธิบายประสิทธิภาพของระบบค้นคืนเอกสารที่เกิดขึ้น โดยในงานวิจัยนี้สนใจการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงหรือความเหมือนระหว่างเอกสารและข้อสอบถามด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) และวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) โดยเลือกใช้ฐานข้อมูลนิตยสารไทม์ (TIME Collection) มาทดสอบเทคนิคการค้นคืนเอกสารที่พัฒนาขึ้น เนื่องจากเป็นนิตยสารข่าว ซึ่งจัดเป็นตัวแทนของชุดเอกสารและข้อสอบถามหนึ่งที่มีความใกล้เคียงกับเอกสารที่ใช้ประกอบการตัดสินใจทางธุรกิจ

#### 6.2 สรุปผลการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อทดสอบประสิทธิภาพของการค้นคืนเอกสารโดยการใช้เทคนิคปริภูมิเวกเตอร์ (Vector Space Model) ด้วยวิธีการวัดความคล้ายคลึง (Similarity) ระหว่างเอกสารและข้อสอบถามที่ต่างกัน โดยกำหนดวัตถุประสงค์การวิจัย ดังนี้

1. ศึกษาการประยุกต์ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) ในการกำหนดกรอบของค่าความคล้ายคลึง สำหรับเทคนิคปริภูมิเวกเตอร์ที่วัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน
2. เปรียบเทียบประสิทธิภาพของการค้นคืนเอกสาร (Document Retrieval) ที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยการวัดความคล้ายคลึงเชิงมุม กับการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

การค้นคืนเอกสารด้วยการใช้เทคนิคปริภูมิเวกเตอร์นั้นเป็นเทคนิคที่เปลี่ยนเอกสารและข้อสอบถามให้อยู่ในรูปแบบเวกเตอร์ ที่แต่ละตำแหน่งมีมิติแทนด้วยค่าน้ำหนักคำที่อยู่ในระบบ โดย

ใช้ทฤษฎีของการค้นคืนสารสนเทศ (Information Retrieval) คือ การให้ค่าน้ำหนักคำ (Term Weighting) ด้วยการพิจารณาจากความถี่ของคำ (Term Frequency) และความถี่ของเอกสารแบบผกผัน (Inverse Document Frequency) และการคำนวณค่าความคล้ายคลึงหรือความเหมือนระหว่างเอกสารและข้อสอบถาม และค้นคืนเอกสารโดยเปรียบเทียบค่าความเหมือนระหว่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถาม แล้วดึงเอกสารที่มีค่าความเหมือนกันภายในค่าที่กำหนดไว้ (threshold) ออกมาแสดงต่อผู้ใช้

ในงานวิจัยนี้สนใจการค้นคืนเอกสารเทคนิคปริภูมิเวกเตอร์ (Vector Space Model) ด้วยวิธีการวัดความคล้ายคลึงหรือความเหมือนระหว่างเอกสารและข้อสอบถามด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) และการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ว่ามีประสิทธิภาพแตกต่างกันหรือไม่ ซึ่งผู้วิจัยสนใจที่จะเพิ่มประสิทธิภาพการค้นคืนเอกสารด้วยเทคนิคของการทำเหมืองข้อมูลเข้ามารวมกับเทคนิคปริภูมิเวกเตอร์ด้วยการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) โดยการนำมาประยุกต์ใช้กำหนดเงื่อนไข (threshold) ในการเลือกเอกสารที่เป็นคำตอบ หรือการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนด ด้วยการประยุกต์ใช้ทฤษฎีเทคนิคการจัดกลุ่มข้อมูล (Clustering) บนระยะห่างเชิงยูคลิเดียน เพื่อกำหนดกรอบค่าความคล้ายคลึงของการค้นคืนด้วยคำรัศมีกลุ่ม (ดังที่กล่าวในหัวข้อ 4.1) โดยกรอบค่าความคล้ายคลึงของการค้นคืนจะเปลี่ยนแปลงไปตามลักษณะของกลุ่มของเอกสารที่มีความใกล้เคียงกับข้อสอบถาม

ในการศึกษาผู้วิจัยกำหนดให้ทดสอบการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึง 2 รูปแบบดังนี้

- 1) การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)
- 2) การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

ประสิทธิภาพของการค้นคืนเอกสารในงานวิจัยนี้จะพิจารณาความถูกต้องและความครอบคลุมในการค้นคืนเอกสารตามความต้องการของผู้ใช้ โดยจะวัดประสิทธิภาพจากค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) เมื่อนำผลประสิทธิภาพของการค้นคืนเอกสารทั้ง 2 รูปแบบมาวิเคราะห์เปรียบเทียบประสิทธิภาพของการค้นคืนเอกสาร การศึกษาสามารถสรุปได้ว่า

เมื่อพิจารณาค่าเฉลี่ยของค่าประสิทธิภาพทั้ง 3 ค่าของทั้ง 83 ข้อสอบถาม การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม มีประสิทธิภาพมากกว่า การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิด โดยนัยด้วยการค้นคืนเอกสารต่อผู้ใช้งานในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) อย่างมีนัยสำคัญทางสถิติ เมื่อพิจารณาจากค่าเฉลี่ยฮาร์โมนิกซึ่งเป็นการคำนวณจากค่าความแม่นยำและค่าความระลึกมาเฉลี่ยกันจะพบว่า การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดเพียงอย่างเดียว มีประสิทธิภาพของค่าเฉลี่ยฮาร์โมนิกน้อยกว่าการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุมเท่ากับ 86.14% ค่าประสิทธิภาพที่แตกต่างกันมาก สรุปได้ว่า การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดเพียงอย่างเดียวสามารถค้นคืนเอกสารที่เกี่ยวข้องเนื่องกับความต้องการออกมาได้น้อย ถึงแม้จำนวนเอกสารที่ถูกค้นคืนออกมาจะมีจำนวนมาก

ผู้วิจัยตั้งข้อสังเกตว่า ปรากฏการณ์ดังกล่าวอาจมีสาเหตุมาจาก วิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิด (Euclidean Distance) ซึ่งจะเป็นการคำนวณความคล้ายคลึงเชิงระยะทางที่เกิดจากการคำนวณวัดระยะทางระหว่าง 2 จุดให้เป็นรากที่สองของผลรวมของผลต่างของข้อมูลในแต่ละมิติยกกำลังสอง ซึ่งการคำนวณจะขึ้นอยู่กับค่าน้ำหนักของค่าในมิติของเวกเตอร์เอกสารและข้อสอบถาม แต่วิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) การคำนวณจะขึ้นอยู่กับค่าน้ำหนักของค่าในตำแหน่งมิติของเวกเตอร์ข้อสอบถามเท่านั้น จากที่ได้กล่าวไปในหัวข้อที่ 4.5 แสดงว่า วิธีระยะห่างเชิงมุมจะคิดค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถามเฉพาะค่าที่ตรงกันระหว่างข้อสอบถามและเอกสารเท่านั้น แต่วิธีระยะห่างเชิงยูคลิดจะคิดค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถามที่ไม่จำกัดเฉพาะค่าที่ตรงกันระหว่างข้อสอบถามและเอกสาร

นอกจากนี้ผู้วิจัยเห็นว่าอาจจะเป็นเพราะเอกสารนิตยสารไทม์ (Time) จากฐานข้อมูลนิตยสารไทม์ (TIME Collection) เป็นเรื่องราวข่าวสารทั่วไป จึงมีเนื้อหาสาระหลากหลายไม่เป็นหมวดหมู่ ทำให้เมื่อนำมากำหนดเป็นเวกเตอร์ ค่าที่ปรากฏในระบบทั้งหมดจึงเป็นค่าที่อาจจะไม่ได้ปรากฏอยู่ในทุกเอกสาร จากการศึกษาพบว่าจำนวนค่าที่ปรากฏในระบบทั้งหมดเท่ากับ 13,813 ค่า จำนวนค่าที่ปรากฏในแต่ละเอกสารเฉลี่ยเท่ากับ 218 ค่า จึงทำให้เวกเตอร์เอกสารมีค่าในตำแหน่งมิติที่ไม่มีค่าปรากฏเท่ากับ 0 เป็นจำนวนมาก นอกจากนี้จำนวนค่าที่ปรากฏในแต่ละข้อสอบถามคำนวณเฉลี่ยได้เท่ากับ 9 ค่า ส่งผลให้ตำแหน่งมิติเวกเตอร์ข้อสอบถามที่ไม่มีค่า หรือมีค่า

เท่ากับ 0 เป็นจำนวนมากเช่นเดียวกัน ดังนั้นเมื่อนำเวกเตอร์เอกสารไปจัดกลุ่มเอกสารด้วยเทคนิค K-means Clustering บนระยะห่างยูคลิเดียน จะแยกความแตกต่างระหว่างแต่ละเอกสารได้ยาก และอาจเป็นไปได้ว่า กลุ่มที่จัดได้เป็นกลุ่มใหญ่ เป็นกลุ่มที่มีค่าในตำแหน่งมิติเป็น 0 เป็นส่วนใหญ่ หรือเป็นกลุ่มของเอกสารที่มีค่าปรากฏในเอกสารน้อย

ในกระบวนการกำหนดกรอบค่าความคล้ายคลึงการค้นคืน เมื่อนำเวกเตอร์ข้อสอบถามทั้ง 83 ข้อ มาพิจารณาว่าอยู่ในกลุ่มใดในกลุ่มของเวกเตอร์เอกสารที่ได้กำหนดไว้แล้วนั้น ผลการทดลองพบว่าเวกเตอร์ข้อสอบถามจะถูกจัดให้อยู่ในกลุ่มที่มีขนาดใหญ่ที่สุดเสมอ (ผลการทดลองแสดงในภาคผนวก ข) ผู้วิจัยตั้งข้อสังเกตว่า อาจจะเนื่องจากเวกเตอร์ข้อสอบถามจะมีค่าในตำแหน่งมิติส่วนใหญ่เป็น 0 ดังนั้นเอกสารที่มีความคล้ายคลึงกับข้อสอบถาม (เอกสารที่มีระยะห่างระหว่างเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสารที่สั้นที่สุด) จึงกลายเป็นเอกสารที่มีค่าในตำแหน่งมิติส่วนใหญ่เป็น 0 เวกเตอร์ข้อสอบถามจึงถูกจัดกลุ่มให้อยู่ในกลุ่มของเอกสารที่มีค่าปรากฏอยู่น้อย และอาจจะไม่มีค่าปรากฏตรงกันของเวกเตอร์เอกสารและข้อสอบถาม เมื่อข้อสอบถามจัดอยู่ในกลุ่มใหญ่ ส่งผลให้กรอบค่าความคล้ายคลึงการค้นคืนเอกสารสำหรับข้อสอบถามนั้น ๆ กว้าง เป็นสาเหตุให้ระบบค้นคืนเอกสารออกมาเป็นจำนวนมาก ส่งผลให้ประสิทธิภาพของค่าความแม่นยำในการค้นคืนต่ำ ซึ่งผิดวัตถุประสงค์ของกระบวนการค้นคืนเอกสาร ที่ระบุว่าระบบค้นคืนเอกสารจะค้นคืนเอกสารที่ผู้ใช้ต้องการในปริมาณที่มากที่สุด และค้นคืนเอกสารที่ผู้ใช้ไม่ต้องการในปริมาณที่น้อยที่สุด ส่งผลให้ประสิทธิภาพการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) ต่ำกว่าการค้นคืนเอกสารด้วยวิธีวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

ดังนั้นผู้วิจัยมีความคิดเห็นว่า วิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน อาจจะไม่เหมาะสมสำหรับนำมาใช้ในกระบวนการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ หรือกล่าวได้ว่าวิธีการวัดความคล้ายคลึงระหว่างเอกสารและข้อสอบถามด้วยวิธีวัดความเหมือนเชิงมุม (Cosine Angle) จะเหมาะสำหรับการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ มากกว่าวิธีการวัดความเหมือนเชิงระยะทาง (Euclidean distance)

ด้วยเหตุนี้ ผู้วิจัยจึงศึกษาเพิ่มเติมโดยการประยุกต์ใช้เทคนิคการจัดกลุ่มข้อมูลแบบ K-mean Clustering บนระยะห่างเชิงมุม (Cosine angle) มากำหนดเงื่อนไข (threshold) ในการเลือกเอกสารที่เป็นคำตอบ หรือการค้นคืนเอกสารภายในกรอบค่าความคล้ายคลึงเชิงมุม ซึ่งผู้วิจัยได้ทำการทดสอบว่าการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึง

เชิงมุม (Cosine Angle) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูลแบบ K-mean Clustering บนระยะห่างเชิงมุมนั้นจะมีประสิทธิภาพการค้นคืนเอกสารทั้ง 3 ค่าแตกต่างกับการไม่ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) มากำหนดกรอบค่าความคล้ายคลึงการค้นคืนหรือไม่

เมื่อวิเคราะห์ผลการทดลองพบว่า การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูลบนระยะห่างเชิงมุม มีประสิทธิภาพค่าความแม่นยำดีกว่าการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม 98.62% แต่มีประสิทธิภาพค่าความระลึกลดลง 24.29% โดยอาจมีเหตุผลจากการที่เอกสารที่ถูกค้นคืนมีความถูกต้องมาก แต่ไม่ได้เป็นเอกสารที่เกี่ยวข้องกับความต้องการทั้งหมด และเมื่อพิจารณาจากค่าเฉลี่ยฮาร์โมนิก การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูลบนระยะห่างเชิงมุม มีประสิทธิภาพค่าเฉลี่ยฮาร์โมนิกดีกว่าการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม เท่ากับ 49.06%

ดังนั้นจากผลการศึกษาเชิงสำรวจในงานวิจัยนี้ ผู้วิจัยจึงสรุปว่า วิธีการวัดความคล้ายคลึงระหว่างเอกสารและข้อสอบถามด้วยวิธีวัดความเหมือนเชิงมุม (Cosine Angle) จะเหมาะสำหรับการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ และจะมีแนวโน้มที่จะมีประสิทธิผลการค้นคืนที่ดีขึ้นเมื่อกำหนดเงื่อนไข (threshold) ในการแสดงเอกสารต่อผู้ใช้ ด้วยกรอบค่าความคล้ายคลึงที่กำหนดจากผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) บนระยะห่างเชิงมุม

### 6.3 การนำงานวิจัยไปประยุกต์ใช้

ในงานวิจัยนี้สามารถใช้เป็นแนวทางในการศึกษาต่อไป หรือนำไปประยุกต์ใช้ในการค้นคืนเอกสารด้านต่าง ๆ เช่น ชุดเอกสารทางธุรกิจ เพื่อสามารถค้นคืนเอกสารทางธุรกิจออกมาให้ตรงกับความต้องการ โดยผู้วิจัยสามารถแบ่งข้อเสนอได้ดังต่อไปนี้

#### 6.3.1 การนำงานวิจัยไปใช้ในเชิงทฤษฎี

ผลการศึกษาแสดงให้เห็นว่า การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) ซึ่งกรอบค่าความคล้ายคลึงเปลี่ยนแปลงไปตามแต่ละกลุ่มของเอกสารที่มีความใกล้เคียงกับข้อสอบถาม ให้ประสิทธิภาพการค้นคืนเอกสารดีกว่าการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุมที่กำหนดกรอบการค้นคืนด้วยค่าที่

คงที่สำหรับทุกข้อสอบถาม ดังนั้นผู้ที่สนใจในด้านการพัฒนาการค้นคืนเอกสารแบบจำลองปริภูมิเวกเตอร์และใช้ฐานข้อมูลนิตยสารไทม์ควรมุ่งเน้นไปในวิธีการวัดความคล้ายคลึงเชิงมุมระหว่างเอกสารและข้อสอบถามมากกว่าวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน

### 6.3.2 การนำงานวิจัยไปใช้ในเชิงประยุกต์

จากการศึกษาเชิงสำรวจ พบว่าการนำเทคนิคการจัดกลุ่มข้อมูล (Clustering) บนระยะห่างเชิงมุม มากำหนดเงื่อนไข (threshold) หรือกรอบค่าความคล้ายคลึงในการแสดงเอกสารผลลัพธ์ต่อผู้ใช้ สามารถเพิ่มประสิทธิภาพค่าเฉลี่ยฮาร์โมนิคและค่าความแม่นยำให้การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ให้ดียิ่งขึ้นได้ แต่จะส่งผลให้ค่าความระลึกลดต่ำลง ดังนั้นในการนำไปประยุกต์ใช้ควรพิจารณาเลือกใช้เทคนิคที่เหมาะสมกับวัตถุประสงค์ของการค้นคืน กล่าวคือ หากต้องการพัฒนาระบบการค้นคืนเอกสารที่มีประสิทธิภาพค่าความระลึกลดสูง ควรพิจารณาใช้การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมแบบที่มีการกำหนดกรอบค่าความคล้ายคลึงคงที่ หากต้องการพัฒนาระบบการค้นคืนเอกสารที่มีประสิทธิภาพค่าความแม่นยำสูง อาจพิจารณาใช้การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูลบนระยะห่างเชิงมุม

## 6.4 ข้อจำกัดของงานวิจัย

การทดลองพัฒนาเครื่องมือการค้นคืนเอกสารในงานวิจัยนี้ มีข้อจำกัดบางประการดังนี้

- 1) เครื่องมือทดสอบเทคนิคการค้นคืนเอกสารในงานวิจัยนี้สร้างขึ้นเพื่อทดสอบกับชุดเอกสารและข้อสอบถามของฐานข้อมูลนิตยสารไทม์ (TIME Collection) ทำให้ผลการทดลองของการค้นคืนเอกสารที่ได้ เป็นผลมาจากการทดสอบด้วยฐานข้อมูลนี้เท่านั้น
- 2) เทคนิคการจัดกลุ่มเอกสาร (Clustering) เพื่อกำหนดเงื่อนไข (threshold) ในการแสดงเอกสารต่อผู้ใช้ ในงานวิจัยนี้ ใช้เทคนิคการจัดกลุ่มข้อมูลแบบ K-mean Clustering
- 3) การทดสอบเครื่องมือการค้นคืนเอกสารในงานวิจัยนี้ ผู้วิจัยกำหนดค่าต่างๆ ในเครื่องมือทดสอบการค้นคืนเอกสาร ดังนี้
  1. กำหนดค่าความเหมือนต่ำสุดของการค้นคืนเอกสารต่อผู้ใช้ ในเทคนิคปริภูมิเวกเตอร์ที่ไม่ได้ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) เข้าร่วม เท่ากับ 0.0439 หรือเท่ากับค่าเฉลี่ย (Mean) บวกค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของค่าความเหมือนของทุกข้อสอบถามกับทุกเอกสาร

2. กำหนดจำนวนกลุ่มของเอกสาร (K) สำหรับการจัดกลุ่มเอกสารนิตยสารไทม์ ของเทคนิคการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน และเทคนิคการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุมที่ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) เข้าร่วม เป็น 20 กลุ่ม
3. กำหนดกรอบค่าความคล้ายคลึงการค้นคืนเอกสารต่อผู้ใช้ด้วยผลลัพธ์ที่ได้จากการจัดกลุ่มเอกสาร ของการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยค่าระยะทางมากที่สุดของเวกเตอร์เอกสารไปยังจุดศูนย์กลางของกลุ่มนั้น ๆ (ค่ารัศมีกลุ่มหารสอง)

### 6.5 แนวทางการศึกษาต่อเนื่อง

จากข้อจำกัดของงานวิจัย ผู้วิจัยจึงได้เสนอแนวทางการศึกษาต่อเนื่อง สำหรับผู้ที่สนใจศึกษาวิจัยการค้นคืนเอกสาร ดังนี้

- 1) สามารถดำเนินการศึกษาชุดเอกสารและข้อสอบถามอื่น ๆ ที่นอกเหนือจากฐานข้อมูลนิตยสารไทม์ (TIME Collection) เพื่อเพิ่มนัยสำคัญทางสถิติของผลการศึกษานิตยสารไทม์
- 2) สามารถใช้เทคนิคในการจัดกลุ่มเอกสาร (Clustering) แบบอื่น ๆ เพื่อทดลองกำหนดเงื่อนไข (threshold) ในการแสดงเอกสารต่อผู้ใช้
- 3) สามารถใช้เงื่อนไขอื่นในการพิจารณาเลือกจำนวนกลุ่มของเอกสาร (K) ที่เหมาะสมสำหรับชุดเอกสารที่ต้องการทดสอบ
- 4) สามารถตั้งค่าในการทดสอบการค้นคืนเอกสาร ตัวอย่างเช่น กำหนดค่าความเหมือนต่ำสุดของการค้นคืนเอกสารผู้ใช้ หรือกำหนดกรอบค่าความคล้ายคลึงการค้นคืนเอกสารผู้ใช้ ด้วยเกณฑ์อื่น ๆ