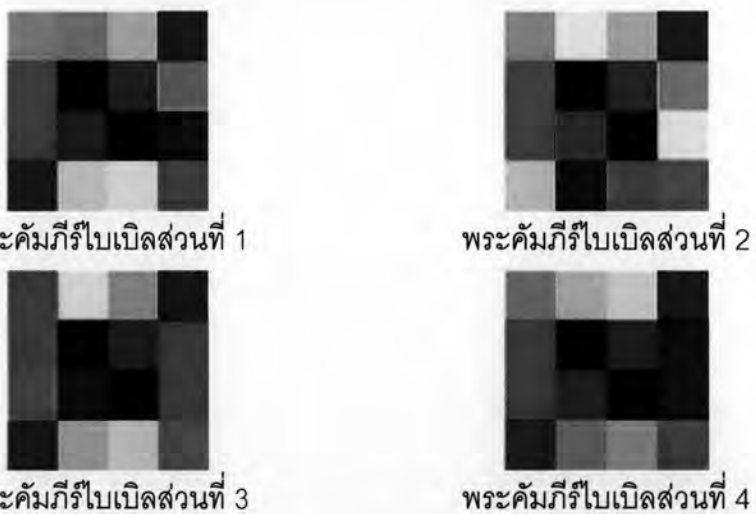
บทที่ 5
ผลภาพบิตแม็บและการทดสอบการแสดงผลภาพสำหรับข้อมูลเอกสารดิจิทัล

ในบทนี้จะนำเสนอผลภาพบิตแม็บจากข้อมูลเอกสารดิจิทัล ทั้งในข้อมูลเอกสารที่มาจากกลุ่มเอกสารประเภทเดียวกัน และต่างประเภทกัน อีกทั้งแสดงแนวทางการพิจารณาภาพบิตแม็บในเบื้องต้น รวมถึงการทดสอบของการแสดงผลภาพสำหรับข้อมูลเอกสารดิจิทัล โดยการพิจารณาภาพบิตแม็บจากอัตโนมัติ จากการจัดกลุ่มด้วยวิธีเคมีน (*k*-Means Clustering) ซึ่งเป็นการจัดกลุ่มข้อมูลจากระยะห่างของภาพบิตแม็บ ทำให้สามารถแยกแยะความเหมือนและความแตกต่างของภาพบิตแม็บได้ชัดเจนและน่าเชื่อถือ มากกว่าการพิจารณาผลภาพบิตแม็บจากข้อมูลเอกสารด้วยการมองด้วยสายตาเพียงอย่างเดียว

5.1 ผลภาพบิตแม็บจากข้อมูลเอกสารประเภทเดียวกัน

เป็นการแสดงผลภาพบิตแม็บที่มาจากเอกสารประเภทเดียวกัน ซึ่งภาพบิตแม็บที่ออกมาควรเป็นภาพที่มีรูปแบบและลักษณะของการแสดงสีที่ใกล้เคียงกัน สามารถพิจารณาจากลักษณะของภาพได้ค่อนข้างชัดเจน ซึ่งแสดงตัวอย่างผลการแสดงภาพบิตแม็บที่มาจากเอกสารในกลุ่มพระคัมภีร์ไบเบิลที่ทำการตัดแบ่งเป็น 4 ส่วน โดยแต่ละส่วนของพระคัมภีร์ไบเบิลไม่มีบทหรือเนื้อหาในเอกสารที่ซ้ำกันเลย ดังรูปที่ 5.1



รูปที่ 5.1 ภาพเอกสารบิตแม็บของเอกสารที่อยู่กลุ่มเดียวกัน

จากรูปที่ 5.1 สังเกตได้ว่าภาพบิตแม็บที่ได้ มีรูปแบบและลักษณะของการแสดงสีที่ใกล้เคียงกันค่อนข้างชัดเจน และมีโครงสร้างของการแสดงสีแต่ละช่องที่เป็นไปในทางเดียวกัน

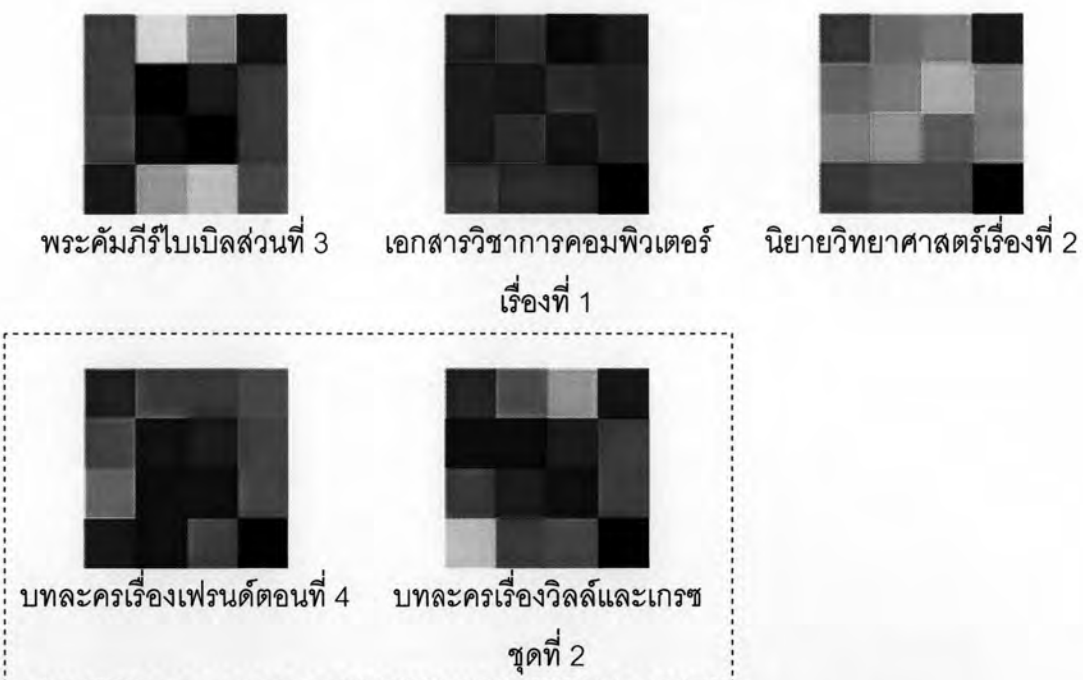
โดยมีการแสดงสีน้ำเงินเข้มและดำบริเวณตรงกลางภาพ และแสดงสีเขียวและตามด้วยสีส้มแดงบริเวณแนวตั้งด้านขวาของภาพ คล้ายกันทั้ง 4 ภาพ

5.2 ผลภาพบิตแม็บจากข้อมูลเอกสารต่างประเภทกัน

เป็นการแสดงผลภาพบิตแม็บที่มาจากเอกสารที่อยู่ต่างประเภทกัน ซึ่งภาพบิตแม็บที่ออกมาควรเป็นภาพที่มีรูปแบบและลักษณะของการแสดงสีที่ต่างกัน สามารถพิจารณาจากลักษณะของภาพที่แตกต่างกันได้ค่อนข้างชัดเจน การแสดงผลภาพบิตแม็บนี้นำเอาภาพบิตแม็บจากเอกสารทั้ง 4 กลุ่มมาทำการเปรียบเทียบ ซึ่งประกอบไปด้วย

- เอกสารพระคัมภีร์ไบเบิล
- เอกสารบทละครโทรทัศน์
- เอกสารนิยายวิทยาศาสตร์
- เอกสารวิชาการทางคอมพิวเตอร์

แสดงตัวอย่างภาพบิตแม็บของเอกสารแต่ละกลุ่ม ดังรูปที่ 5.2



รูปที่ 5.2 ภาพเอกสารบิตแม็บของเอกสารที่อยู่ต่างกลุ่มกัน

รูปที่ 5.2 แสดงภาพบิตแม็บของเอกสารแต่ละกลุ่มที่งานวิจัยนี้นำมาพิจารณา สังเกตได้ว่าภาพบิตแม็บของเอกสารที่อยู่ในกลุ่มบทละคร พระคัมภีร์ไบเบิล และนิยายวิทยาศาสตร์ มีรูปแบบและลักษณะของการแสดงสีที่แตกต่างกันอย่างชัดเจน โดยเฉพาะเอกสารวิชาการคอมพิวเตอร์ที่มี

พื้นที่ส่วนใหญ่ในภาพเป็นสีแดงส้ม และนิยายวิทยาศาสตร์ที่มีการแสดงสีของสีเขียวเหลืองที่เด่นชัด แต่ถ้าพิจารณาภาพบิตแม็บของเอกสารเฉพาะในกลุ่มบทละคร คือ บทละครเรื่องเฟรนด์ และบทละครเรื่องวิลล์และเกรซ พบว่ามีรูปแบบและลักษณะของการแสดงสีที่คล้ายกันมาก แม้เป็นบทละครที่ต่างเรื่องกัน โดยแสดงสีน้ำเงินตัดกับสีแดงส้มตามแนวทแยงของภาพ และแสดงสีเขียวเหลืองตามขอบของภาพบิตแม็บ

จากผลการแสดงภาพบิตแม็บในหัวข้อ 5.1 และ 5.2 แสดงให้เห็นว่าการแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล สามารถช่วยให้การเปรียบเทียบ แยกแยะ และจำแนกเอกสารโดยการพิจารณาภาพบิตแม็บด้วยสายตาได้เป็นอย่างดี (แสดงผลภาพบิตแม็บของเอกสารทั้งหมดในภาคผนวก ง) อย่างไรก็ตามเพื่อเป็นการเพิ่มความน่าเชื่อถือของผลการทดลองมากยิ่งขึ้นและสามารถวัดผลได้ งานวิจัยนี้จึงทำการทดลองจัดกลุ่มเอกสารจากข้อมูลภาพบิตแม็บด้วยวิธีเคมีน ซึ่งเป็นวิธีการที่สามารถวิเคราะห์ถึงคุณสมบัติจากระยะทางระหว่างภาพบิตแม็บ ที่สามารถบอกถึงความแตกต่างของข้อมูลเอกสารได้

5.3 การพิจารณาภาพบิตแม็บโดยการจัดกลุ่มด้วยวิธีเคมีน

เป็นการพิจารณาเปรียบเทียบความเหมือนและความแตกต่างของภาพบิตแม็บ โดยอาศัยการจัดกลุ่มด้วยวิธีเคมีน ซึ่งเป็นจำแนกประเภทของข้อมูลที่มีความคล้ายกันให้อยู่ในกลุ่มเดียวกันและจัดแยกข้อมูลที่แตกต่างกันให้อยู่คนละกลุ่มกัน โดยการวิเคราะห์จากระยะทางของข้อมูลด้วยการคำนวณระยะทางแบบแมนฮัทตัน (Manhattan Distance) ซึ่งแสดงรายละเอียดการจัดกลุ่มด้วยวิธีเคมีนในหัวข้อ 2.4

งานวิจัยนี้เลือกการจัดกลุ่มโดยวิธีเคมีน เนื่องจากเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพประมวลผลเร็ว เหมาะสมกับการพิจารณาข้อมูลที่ทราบจำนวนกลุ่มก่อนหน้า และผลจากการจัดกลุ่มภาพบิตแม็บของเอกสารโดยวิธีเคมีน เป็นวิธีที่ใช้ในการวัดผลของการแสดงผลภาพบิตแม็บจากเอกสารดิจิทัลได้เป็นอย่างดี เพราะเป็นวิธีที่วิเคราะห์ถึงคุณสมบัติจากระยะทางระหว่างภาพบิตแม็บ ทำให้สามารถพิจารณาเปรียบเทียบความเหมือนและความแตกต่างของภาพบิตแม็บได้ชัดเจนมากขึ้น

การทดสอบการจัดกลุ่มด้วยวิธีเคมีนของภาพบิตแม็บจากข้อมูลเอกสาร จะใช้เอกสารทุกเอกสารมาทำการทดสอบ ซึ่งมีเอกสารทั้งหมดจำนวน 22 เอกสาร แบ่งออกได้เป็น 4 กลุ่ม โดยใช้ค่าความถี่ของการนับอักขระที่กำกับอยู่ในตารางเมทริกซ์ มาทำการคำนวณค่าระยะทางระหว่างเอกสารด้วยวิธีแบบแมนฮัทตัน (แสดงระยะทางระหว่างคู่เอกสารทั้งหมดใน ภาคผนวก ค) และทำการทดสอบการจัดกลุ่มของภาพบิตแม็บด้วยวิธีเคมีนจำนวน 20 ครั้ง

ผลการทดสอบการจัดกลุ่มในเบื้องต้นพบว่า สามารถทำการจัดกลุ่มได้ถูกต้องเฉลี่ย 78.18 เปอร์เซ็นต์ ซึ่งสรุปได้ว่าการแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล ยังทำการจัดกลุ่มเอกสารผิดพลาดอยู่บ้าง เนื่องจากได้สมาชิกกลุ่มที่ไม่แน่นอนจากการจัดกลุ่มในแต่ละครั้ง

จากการศึกษาเพื่อหาสาเหตุดังกล่าวพบว่า ภาพบิตแม็บของเอกสารมีระยะทางที่แตกต่างกันไม่มาก และจำนวนเอกสารที่นำมาทดลองมีจำนวนน้อย ซึ่งจากการศึกษาเพิ่มเติมพบว่าสาเหตุดังกล่าวเป็นข้อจำกัดของวิธีเคมีน เนื่องจากวิธีเคมีนจะทำการเลือกตัวแทนกลุ่มเริ่มต้นจากข้อมูลทั้งหมดโดยการสุ่ม หากข้อมูลมีความแตกต่างของระยะทางไม่มาก และมีจำนวนข้อมูลน้อย ส่งผลให้การเลือกตัวแทนกลุ่มเริ่มต้นอาจจะอยู่ในกลุ่มข้อมูลเดียวกัน ทำให้ผลของการจัดกลุ่มเกิดความผิดพลาด

เพื่อทำให้การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล สามารถทำการจัดกลุ่มเอกสารให้มีความถูกต้องมากยิ่งขึ้น งานวิจัยนี้จึงทำการปรับอัลกอริธึมของเคมีนเพิ่มเติม ในขั้นตอนของการเลือกตัวแทนกลุ่มเริ่มต้น โดยทำการตรวจสอบระยะห่างของตัวแทนกลุ่มเริ่มต้นในแต่ละกลุ่ม ให้มีความแตกต่างกันไม่น้อยกว่าค่าระยะห่างที่มากที่สุดของเอกสารในกลุ่มเดียวกัน ทำให้การเลือกตัวแทนกลุ่มเริ่มต้นสามารถทำได้ดีและมีประสิทธิภาพมากยิ่งขึ้น สามารถป้องกันไม่ให้เกิดการเลือกตัวแทนกลุ่มเริ่มต้นที่ไม่ดี นอกจากนี้ยังเพิ่มขั้นตอนการเปรียบเทียบสมาชิกที่ได้จากการจัดกลุ่มที่ต้องได้สมาชิกในกลุ่มเหมือนกันอย่างน้อย 2 ครั้ง

ผลการทดสอบการจัดกลุ่มหลังทำการปรับอัลกอริธึมของเคมีนเพิ่มเติม พบว่าสามารถทำการจัดกลุ่มได้ถูกต้องเฉลี่ย 92.72 เปอร์เซ็นต์ โดยให้ผลการจัดกลุ่มที่ถูกต้องเพิ่มขึ้นจากเดิม 14.54 เปอร์เซ็นต์ ซึ่งสรุปรายละเอียดของผลการทดสอบการจัดกลุ่มจากเอกสารทั้ง 22 เอกสารด้วยอัลกอริธึมของเคมีน โดยแสดงผลการจัดกลุ่มเฉลี่ยจากการจัดกลุ่มจำนวน 20 ครั้ง ดังตารางที่ 5.1

ตารางที่ 5.1 สรุปผลการทดสอบการจัดกลุ่มเอกสารด้วยวิธีเคมีน

ประเภทการจัดกลุ่ม	จำนวนเอกสารที่จัดกลุ่มถูกต้อง (เฉลี่ย)	เปอร์เซ็นต์ความถูกต้อง (เฉลี่ย)
การจัดกลุ่มในเบื้องต้น	17.2	78.18
การจัดกลุ่มหลังทำการปรับอัลกอริธึมของเคมีนเพิ่มเติม	20.4	92.72
เพิ่มขึ้น		14.54

การปรับอัลกอริธึมของวิธีเคมีน นอกจากจะทำให้ผลการจัดกลุ่มข้อมูลเอกสารที่มีจำนวนน้อย และมีระยะห่างระหว่างเอกสารไม่มาก มีความถูกต้องเพิ่มขึ้นแล้ว ยังสามารถนำปรับไปใช้ในการจัดกลุ่มข้อมูลเอกสารในชีวิตประจำวันได้อีกด้วย เนื่องจากจำนวนของเอกสารของแต่ละคนโดยปกติมักมีจำนวนไม่มาก จึงถือเป็นประโยชน์ที่การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัลสามารถที่จะทำการจัดกลุ่มกับเอกสารเหล่านี้ได้ถูกต้องมากยิ่งขึ้น แสดงอัลกอริธึมของวิธีการการจัดกลุ่มแบบเคมีนที่ทำการปรับเพิ่มเติม ดังรูปที่ 5.3

อัลกอริทึมการจัดกลุ่มด้วยวิธีเคมีนที่ทำการปรับเพิ่ม

ข้อมูลเริ่มต้น: วัตถุที่นำมาจัดกลุ่ม ระยะห่างของตัวแทนกลุ่มเริ่มต้น d และจำนวนกลุ่ม *k*

ผลลัพธ์: ชุดของกลุ่มจำนวน *k* กลุ่ม

วิธีการ:

(1) ทำซ้ำ

(1.1) ทำการเลือกวัตถุจำนวน *k* ตัว เพื่อมาเป็นตัวแทนกลุ่มเริ่มต้น

(1.2) ทำการตรวจสอบระยะห่างของตัวแทนกลุ่มเริ่มต้น

(1.3) เริ่มทำ (1.1) อีกครั้ง จนกระทั่งระยะห่างของตัวแทนกลุ่มเริ่มต้นไม่น้อยกว่า *d*

(2) ทำซ้ำ

(2.1) ทำการเลือกวัตถุในกลุ่ม จากระยะทางของวัตถุกับตัวแทนกลุ่ม

(2.2) เริ่มทำ (2.1) อีกครั้ง จนกระทั่งสมาชิกในกลุ่มไม่มีการเปลี่ยนแปลง

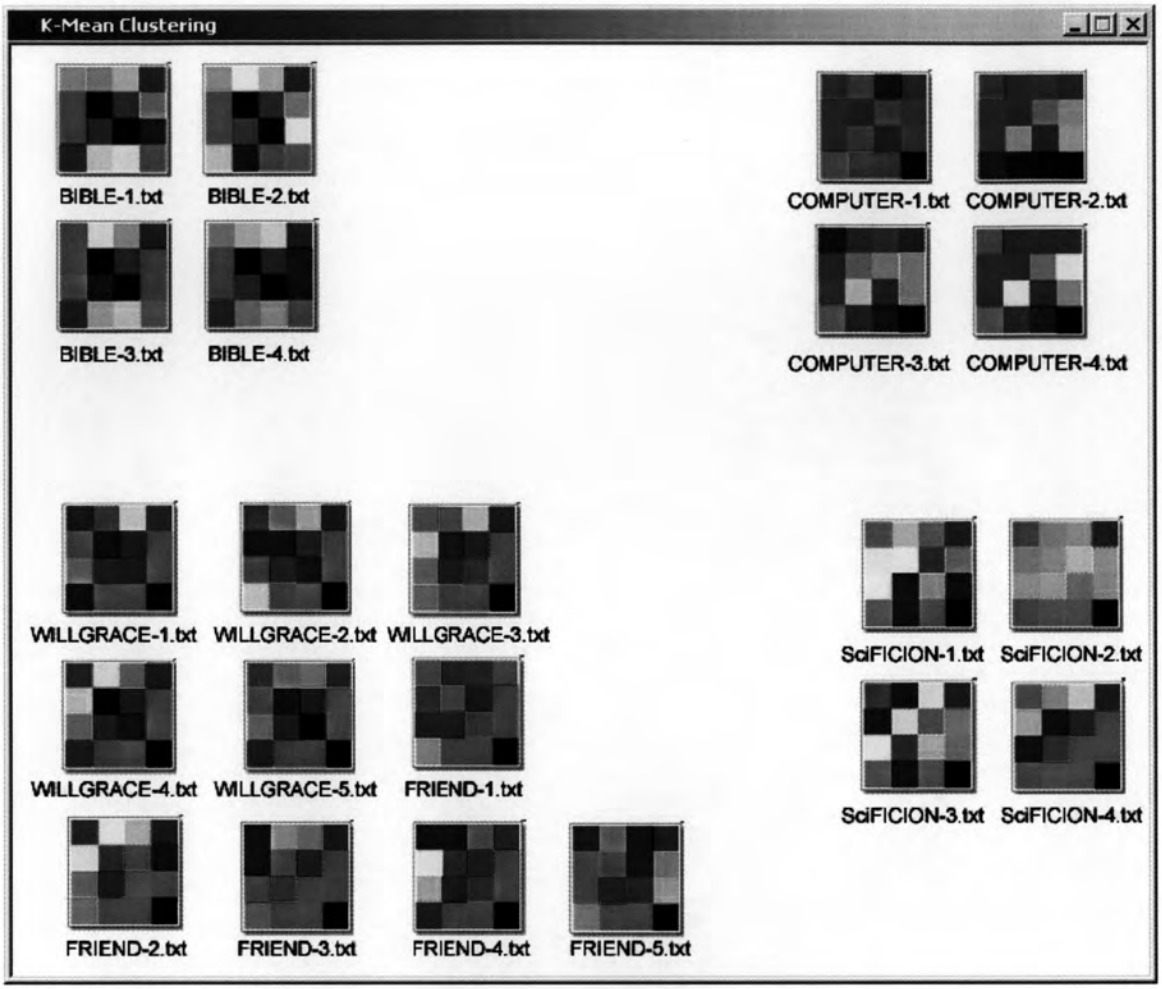
(4) เก็บข้อมูลสมาชิกในกลุ่มทุกกลุ่ม

(5) เริ่มทำ (1) อีกครั้ง จนกระทั่งมีการเก็บข้อมูลสมาชิกในกลุ่มทุกกลุ่มจำนวน 2 ครั้ง

(6) เริ่มทำ (1) อีกครั้ง จนกระทั่งข้อมูลสมาชิกในกลุ่มจากการจัดกลุ่มทั้ง 2 ครั้งไม่แตกต่างกัน

รูปที่ 5.3 อัลกอริธึมของวิธีการการจัดกลุ่มแบบเคมีนที่ทำการปรับเพิ่มเติม

จากผลการทดลองสามารถสรุปได้ว่า การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล สามารถทำการวิเคราะห์ข้อมูลเอกสารได้ตรงตามคุณลักษณะของเอกสาร ซึ่งสามารถแยกแยะประเภทเอกสารได้โดยการจัดกลุ่มแบบวิธีเคมีน แสดงผลการจัดกลุ่มของเอกสารทั้งหมดดังรูปที่ 5.4



รูปที่ 5.4 ผลการจัดกลุ่มของภาพเอกสารบิตแม็บด้วยวิธีเคมีน