

บทที่ 5 บทสรุป

ในบทนี้จะกล่าวถึง บทสรุป ปัญหาและอุปสรรคในการวิจัย รวมถึงข้อเสนอแนะในการ
ทำวิจัย โดยมีรายละเอียดดังต่อไปนี้

5.1 บทสรุป

การวิจัยนี้นำเสนอการรู้จำตัวพิมพ์อักษรธรรมล้านนา โดยนำข้อมูลภาพเอกสาร
ตัวพิมพ์อักษรธรรมล้านนาจากโครงการ e-60 วรรณพิมพ์ล้านนา ซึ่งเป็นโครงการที่รวบรวม
วรรณกรรมภาษาล้านนาสำคัญในหมวดต่างๆ โดยแบ่งกระบวนการออกเป็นสองส่วน ในส่วนแรก
เป็นการออกแบบกระบวนการรู้จำตัวพิมพ์อักษรธรรมล้านนา ซึ่งประกอบด้วยหกขั้นตอน ขั้นตอน
ที่หนึ่ง การประมวลผลภาพเบื้องต้น โดยใช้เทคนิคการแปลงภาพให้เป็นสองระดับด้วยวิธีของโอที
จากนั้นกำจัดสัญญาณรบกวนด้วยการดำเนินการเชิงสัญญาณวิทยาและการระบุป้ายให้องค์ประกอบ
ที่เชื่อมต่อ และหาความเอียงภาพด้วยเทคนิคการปรับให้เรียบด้วยการนับความยาว เพื่อใช้ปรับภาพ
เอกสารให้ตรงและง่ายต่อการตัดแบ่ง ขั้นตอนที่สอง การตัดแบ่ง ประกอบด้วยสามขั้นตอนย่อย คือ
การตัดแบ่งบรรทัด การตัดแบ่งบล็อก และการหาจุดแบ่งระดับของตัวอักษร โดยการตัดแบ่งเป็น
บรรทัดใช้เทคนิคโพเรจชัน โพรไฟล์ในแนวนอน การตัดแบ่งบล็อกใช้เทคนิคโพเรจชัน โพรไฟล์
ในแนวตั้ง และการหาจุดแบ่งระดับของตัวอักษรด้วยการ โพเรจชัน โพรไฟล์ในแนวนอนของภาพ
เอกสารแต่ละบรรทัด ร่วมกับการปรับฮิสโตแกรมให้เรียบขึ้นด้วยค่าเกาส์เซียน ขั้นตอนที่สาม
การสร้างชุดข้อมูลฝึกสอนด้วยการจัดกลุ่มแบบเคมีน เพื่อใช้ในการฝึกสอนตัวจำแนกประเภท
ขั้นตอนที่สี่ การจำแนกประเภทตัวอักษรด้วยเค-เนียร์เรสเนเบอร์ เป็นวิธีพิจารณาระยะห่างระหว่าง
บล็อกตัวอักษรที่สนใจกับตัวอักษรที่อยู่ในชุดข้อมูลฝึกสอน ขั้นตอนที่ห้า การจำแนกประเภท
ตัวอักษรด้วยเทมเพลตแมทซิง หากระยะห่างระหว่างบล็อกตัวอักษรที่สนใจกับตัวอักษรที่อยู่ในชุด
ข้อมูลฝึกสอนมากเกินไปที่ยอมรับได้ บล็อกจะถูกจำแนกซ้ำด้วยเทมเพลตแมทซิง ขั้นตอนที่หก
การจำแนกประเภทตัวอักษรด้วยคอนดิชันนัลแรนคอมฟิลด์ เป็นขั้นตอนสุดท้ายของการจำแนก
เพื่อตรวจสอบความถูกต้องของผลลัพธ์ด้วยการพิจารณาตัวอักษรข้างเคียง

ในส่วนที่สองเป็นขั้นตอนในการทดลองเพื่อวัดประสิทธิภาพกระบวนการรู้จำ โดยแบ่ง
การทดลองออกเป็นสามขั้นตอน ขั้นตอนที่หนึ่ง การทดสอบประสิทธิภาพการแบ่งบรรทัด พบว่า
สามารถตัดแบ่งบรรทัดได้ถูกต้องถึงร้อยละ 97.14 แต่ส่วนที่ตัดแบ่งผิดพลาดพบว่า เกิดจากการ
ภาพอักษรเอียงบางบรรทัด ไม่ได้เอียงทั้งหน้าเอกสาร อีกทั้งมีตัวอักษรเหลื่อมกันระหว่างบรรทัด
ทำให้ไม่สามารถแยกบรรทัดออกจากกันได้ ขั้นตอนที่สอง การทดสอบประสิทธิภาพการตัดแบ่ง

บล็อกตัวอักษร จากผลการทดลองพบว่าสามารถตัดแบ่งบล็อกได้ถูกต้องถึงร้อยละ 98.63 และขั้นตอนที่สาม การทดสอบประสิทธิภาพการรู้จำตัวอักษร มีชุดข้อมูลทดสอบทั้งหมด 4,722 บล็อก หรือ 9,656 ตัวอักษร ได้แบ่งเป็น 4 การทดสอบ คือ การจำแนกด้วยซอฟต์แวร์เวกเตอร์แมชชีน การจำแนกด้วยเค-เนียร์เรสเนเบอร์ การจำแนกด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทซิง และการจำแนกด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทซิงและคอนดิชันนัลแรนคอมฟิลด์ จากผลการทดลองพบว่าการทดสอบที่สี่ การจำแนกด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทซิงและคอนดิชันนัลแรนคอมฟิลด์ เมื่อใช้ 2 คุณลักษณะ คือ หมวดของอักษรธรรมล้านนา และระดับของตัวอักษร และใช้รูปแบบการพิจารณาตัวอักษรข้างเคียงแบบหน้า 2 หลัง 2 สำหรับการจำแนกด้วยคอนดิชันนัลแรนคอมฟิลด์ มีประสิทธิภาพสูงที่สุดในการรู้จำตัวพิมพ์อักษรธรรมล้านนา โดยมีค่า F-Measure สูงถึง 0.93928 หรือร้อยละ 93.93 แต่อย่างไรก็ตาม สำหรับการจำแนกด้วยเค-เนียร์เรสเนเบอร์ร่วมกับเทมเพลตแมทซิง ตัวอักษรที่สามารถจำแนกได้ถูกต้องนั้น จะต้องมียุคหลายเส้นที่ค่อนข้างสมบูรณ์ ขาดหายเพียงเล็กน้อย สำหรับการจำแนกด้วยคอนดิชันนัลแรนคอมฟิลด์ ตัวอักษรที่สามารถจำแนกได้ถูกต้องนั้นมักเป็นอักษรที่มักจำแนกผิดพลาดจากการจำแนกด้วยเค-เนียร์เรสเนเบอร์และเทมเพลตแมทซิง และเกิดร่วมกับอักษรข้างเคียงตัวเดิมเป็นประจำ

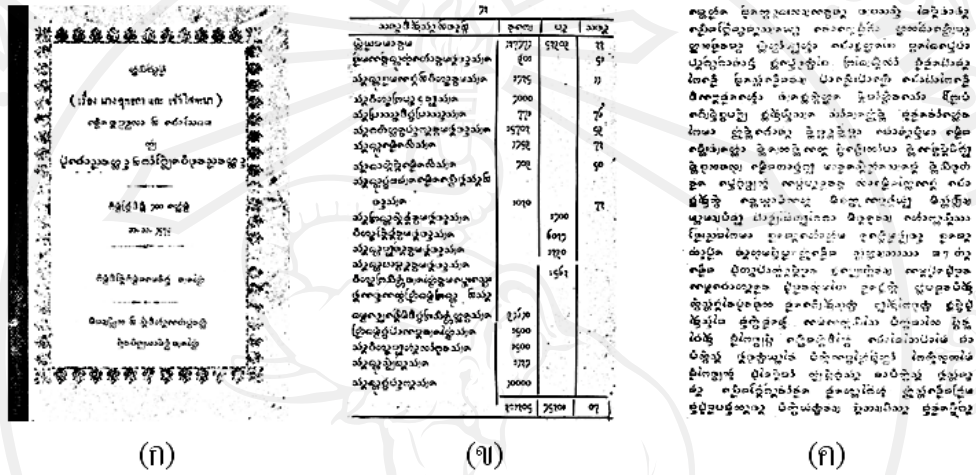
5.2 ปัญหาและอุปสรรคในการวิจัย

ลักษณะของภาพเอกสารภาษาล้านนาที่ใช้งานวิจัยนี้ พบว่ามี 3 ลักษณะคือ ภาพเอกสารที่มีเฉพาะส่วนของข้อความดังรูป 5.1 (ก) ภาพเอกสารที่มีทั้งรูปภาพและข้อความประกอบกัน ดังรูป 5.1 (ข) และภาพเอกสารที่มีทั้งข้อความ ตาราง หรือกรอบข้อความประกอบกัน ดังรูป 5.1 (ค) ซึ่งในการวิจัยนี้ได้เลือกภาพเอกสารที่มีเฉพาะข้อความเท่านั้นมาใช้ในการทดสอบ และต้องเป็นภาพที่มีความกว้างอยู่ในช่วง 1,000 - 1,450 พิกเซล และมีความสูงอยู่ในช่วง 1,600 - 2,200 พิกเซล เนื่องจากการทดสอบด้วยเทมเพลตแมทซิง ใช้ภาพอักษรขนาดที่ไม่ถูกปรับเป็นมาตรฐาน หากภาพเอกสารมีขนาดแตกต่างจากข้อกำหนดมากอาจทำให้อัตราความถูกต้องในการรู้จำตัวอักษรลดลง

5.3 ข้อเสนอแนะในการทำวิจัย

ในขั้นตอนการจำแนกด้วยเค-เนียร์เรสเนเบอร์ อาจเปลี่ยนตัวจำแนกประเภท หรือคุณลักษณะอื่นที่มีผลทำให้การรู้จำมีประสิทธิภาพมากขึ้น แล้วจึงใช้ร่วมกับการจำแนกด้วยเทมเพลตแมทซิงและคอนดิชันนัลแรนคอมฟิลด์ เพื่อตรวจสอบความถูกต้อง และจากข้อกำหนดในการเลือกภาพเอกสารที่ใช้ในการวิจัยนี้ ทำให้วิธีการรู้จำตัวพิมพ์อักษรธรรมล้านนาที่นำเสนอในงานวิจัยยังไม่สามารถใช้กับภาพเอกสารภาษาล้านนาทุกรูปแบบได้ เนื่องจากลักษณะของภาพ

เอกสารแต่ละแบบมีข้อจำกัดและปัญหาที่ต้องแก้ไขก่อนนำไปใช้งาน ดังนั้นในการนำแนวคิดไปพัฒนาต่อในอนาคต ควรให้ความสำคัญต่อรายละเอียดเกี่ยวกับข้อจำกัดของภาพเอกสารดังที่กล่าวไปแล้วข้างต้นด้วย หากสามารถลดข้อจำกัดเกี่ยวกับภาพเอกสารที่ใช้ในการวิจัยได้ จะทำให้รู้จำได้ถูกต้องและสมบูรณ์มากยิ่งขึ้น



รูป 5.1 ลักษณะของภาพเอกสารตัวพิมพ์อักษรธรรมล้านนา