

บทที่ 5

สรุปผลการทดลอง ข้อเสนอแนะ

สรุปผลการทดลอง

การตัดแยกตัวอักษรภาษาไทยที่ติดกัน มีขึ้นเพื่อช่วยให้การรู้จำตัวอักษรภาษาไทยมีความถูกต้องเพิ่มมากขึ้น โดยในสัณนิเคราะห์จะแบ่งตัวอักษรออกเป็นกลุ่มเพื่อให้ง่ายต่อการวิเคราะห์ว่ามี การติดกันในลักษณะใด เมื่อตรวจสอบได้ว่าอยู่กลุ่มใดแล้ว จึงทำการตัดแยกโดย ส่วนตัดแยก จากงานวิจัยของวัชร วัชรวิริยะ [5] ทำการตัดแยกตัวอักษรจากบรรทัด และคอลัมน์ โดยใช้โปรเจกชันในแนวนอน และแนวตั้งเท่านั้นไม่ได้ทำการวิเคราะห์ และตัดแยกส่วนที่ติดกันก่อนส่วนไปยังส่วนรู้จำ ส่วนงานวิจัยของ Nucharee Premchaiswadi, Wichian Premchaiswadi, Seinosuke Narita [7] นั้นทำการแบ่งลักษณะตัวอักษรออกเป็น 3 กลุ่ม คือ 1. พยัญชนะ 2. สระ 3. วรณยุกต์ ซึ่งสามารถแบ่งเป็นกลุ่มตัวอักษรที่ติดกันได้ 6 กลุ่มด้วยกัน โดยการหาแนวตัดแยกแบ่งเป็น 2 แบบคือ 1. ใช้ค่าที่น้อยที่สุดของโปรเจกชันทั้งในแนวตั้ง และแนวนอน ในการตัดแยก 2. ใช้อัตราส่วนระหว่างอนุพันธ์อันดับสองของค่าโปรเจกชัน กับค่าโปรเจกชันในการตัดแยก (Kahan and Pavlidis [11]) และการตัดแยกใช้วิธีตัดแบบตรงๆ ซึ่งมีผลให้ส่วนของตัวอักษรที่ทับกันเสียรูปแบบไป และไม่ได้สนใจตัวอักษรที่ไขว้กัน ส่วนงานวิจัยของวิชา พานิช [1] มีความใกล้เคียงกับ Nucharee Premchiswadi มากเพียงแต่เพิ่มในส่วนของการแก้ไขความเอียงของเอกสารที่ทำการสแกนเข้ามา ซึ่งค่าความเอียงที่ยอมรับได้คือไม่เกิน 5 องศา และส่วนตัดแยกของระบบที่เพิ่มขึ้น การตัดของตัวอักษรที่ติดกันแบบซ้อนทับกัน เช่น ทีไ จะใช้วิธีการตัดแบบตรง ทำให้ส่วนของไม้มีขนาดหายไป และยังไม่มีการตัดในส่วนของตัวอักษรที่ไขว้กัน เช่น บั โดยจะส่งไปยังส่วนรู้จำทั้งที่ยังติดกัน

จากการทดสอบการตัดแยกตัวอักษรไทยที่ติดกันพบว่า ได้อัตราความถูกต้องเฉลี่ย 63.14% จากตัวอักษรที่ติดกัน และเวลาในการตัดแยกเอกสารทั้งหมดที่ 22.07 ตัวอักษรต่อวินาที โดยใช้แบบตัวอักษรดังนี้ AngsanaUPC, BrowalliaUPC, CordiaUPC, DilleniaUPC, EucrosiaUPC, FreesiaUPC, IrisUPC, JasmineUPC และ KodchiangUPC ที่ขนาดตัวอักษร 12-36 points โดยตัวอักษรขนาด 12 point มีอัตราความถูกต้องเฉลี่ยทุกแบบตัวอักษรที่ติดกัน 54.75% เวลาเฉลี่ยทั้งหมด 68.53 ตัวอักษรต่อวินาที ส่วนตัวอักษรขนาด 28 point มีอัตราความถูกต้องเฉลี่ยทุกแบบตัวอักษรที่ติดกัน 68.51% เวลาเฉลี่ยทั้งหมด 14.47 ตัวอักษรต่อวินาที จึงเห็น

ได้ว่าเมื่อตัวอักษรที่มีขนาดใหญ่ขึ้นเปอร์เซ็นต์ความถูกต้องในการตัดก็เพิ่มขึ้น แต่จะใช้เวลาในการตัดเพิ่มขึ้น ทั้งนี้เนื่องจากตัวอักษรที่มีขนาดใหญ่ใช้พื้นที่มากจึงมีจำนวนจุดของภาพมากกว่านั่นเอง

ความผิดพลาดที่เกิดขึ้นกับทุกส่วนของระบบการตัดแยก โดยระบบตัดแยกคอลัมน์ผิดพลาด ก็จะทำให้เกิดความผิดพลาดในการเรียงลำดับ ถ้าระบบตัดบรรทัดผิดพลาดจะเป็นการผิดพลาดต่อกันทั้งบรรทัด ถ้าอักษรที่สแกนมีขนาดเล็ก ทำให้มีตัวอักษรที่ติดกันมาก ก็จะทำให้เกิดความผิดพลาดในส่วนตัดแยกเพิ่มขึ้นมา

ข้อเสนอแนะ

แนวทางในการพัฒนาที่น่าจะทำต่อไปคือ สามารถนำไปใช้ในการตัดแยกตัวอักษรภาษาไทย ที่เป็นลายมือ หรือรูปแบบตัวอักษรภาษาไทยอื่นๆ ที่หลากหลายมากขึ้น และการตัดแยกส่วนของตัวอักษรที่มีลักษณะ เอียง ตัวหนา หรือแม้กระทั่ง ตัวอักษรที่มีหลายๆ ขนาดในบรรทัดเดียวกัน และเวลาในการตัดแยกตัวอักษรภาษาไทยควรพัฒนาในต่ำลงไปโดยใช้เทคนิคทางโปรแกรมมิ่ง เพื่อสามารถนำไปประยุกต์ใช้งานให้รวดเร็วยิ่งขึ้น