



สรุปการวิจัยและข้อเสนอแนะ

4.1 สรุปการวิจัย

การศึกษาการรู้จำสายอักขระตัวพิมพ์ไทยโดยวิธีซินแทกติกนั้น เป็นการศึกษาถึงความเป็นไปได้ของการนำเอาภาพของตัวอักษรมาทำการรู้จำซึ่งมีขั้นตอนหลัก ๆ อยู่ 3 ขั้นตอนคือ ขั้นตอนการแยกข้อมูลภาพออกมาเป็นกลุ่มของข้อมูลภาพของตัวอักษรแต่ละตัว ขั้นตอนการจัดเรียงตัวอักษร ขั้นตอนการรู้จำตัวอักษร โดยในขั้นตอนแรกจะทำการอ่านภาพเข้ามาและจัดเก็บข้อมูลภาพที่ต้องการรู้จำในลักษณะที่สะดวกต่อการนำไปใช้งานต่อไป จากนั้นก็นำข้อมูลภาพที่จัดเก็บไว้มาทำการแยกออกมาเป็นกลุ่มข้อมูลภาพของตัวอักษรแต่ละตัว ซึ่งในขั้นตอนนี้หากพบว่าตัวอักษรติดกัน ก็จำเป็นต้องทำการเปรียบเทียบกับตัวอักษรต้นแบบสำหรับในกรณีที่มีกลุ่มข้อมูลภาพของตัวอักษรที่ติดกันอยู่ ผลลัพธ์ที่ได้จากขั้นตอนนี้จะถูกนำไปทำการจัดเรียงตัวอักษรใหม่ ให้ถูกต้องตามหลักการพิมพ์ของ word processor เมื่อทำการจัดเรียงเรียบร้อยแล้ว ก็จะนำไปทำการรู้จำซึ่งในการรู้จำได้ใช้วิธีการในการจำแนกตัวอักษรโดยการหาค่าระยะระหว่างต้นไม้ของรูปสัญลักษณ์ กับการเปรียบเทียบตัวอักษรทางกายภาพระหว่างตัวอักษรต้นแบบและตัวอักษรที่ต้องการรู้จำ ซึ่งการทดสอบการรู้จำของระบบ จะใช้ตัวอักษรที่ทดสอบเป็นตัวอักษรแบบตัวตรง ลักษณะตัวอักษรแบบ EucrosiaUPC ขนาด 18 point ผลการรู้จำจากตัวอย่างสายอักขระทั้งหมด 150 สาย รวมเป็นตัวอักษรทั้งสิ้น 1974 ตัวอักษร พบว่ามีอัตราการรู้จำที่ถูกต้อง 92.70 % การรู้จำผิด 4.40 % ส่วนมากเกิดขึ้นกับตัวอักษรที่มีลักษณะใกล้เคียงกันและมีขนาดเล็ก และรู้จำไม่ได้ 2.90 % ส่วนมากเกิดขึ้นกับตัวอักษรที่ติดกันที่ไม่สามารถแยกออกได้

จากระบบที่พัฒนาขึ้นนั้น ข้อมูลภาพจะถูกเก็บลงเพิ่มข้อมูล ซึ่งมีอยู่ด้วยกัน 2 เพิ่มข้อมูล คือ เพิ่มข้อมูลที่เป็นข้อมูลภาพของเอกสารที่ต้องการทดสอบที่ได้ทำการปรับปรุงภาพแล้ว และเพิ่มข้อมูลที่เป็นตำแหน่งของตัวอักษรแต่ละตัวว่าอยู่ที่ใด มีลักษณะเป็นตัวอักษรประเภทใด มีการรบกวนจากตัวอักษรข้างเคียงหรือไม่ ซึ่งจะใช้เนื้อที่ในการเก็บข้อมูลของตัวอักษรแต่ละตัว 10 ไบต์ ในกรณีที่ไม่มีการรบกวนจากตัวอักษรข้างเคียง สำหรับกรณีที่มีการรบกวนจากตัวอักษรข้างเคียงข้อมูลของตัวอักษรก็จะแตกต่างกัน โดยขึ้นอยู่กับจำนวนของกลุ่มข้อมูลของตัวอักษรข้างเคียงว่ามีเท่าไร โดยจำนวนข้อมูลต่อตัวอักษรแต่ละตัว ในกรณีที่มีจำนวนกลุ่มข้อมูลเป็น 1 2 และ 3 จำนวนข้อมูลของตัวอักษรจะเป็น 18 26 และ 34 ตามลำดับและขนาดของภาพต้นแบบที่ใช้ในการทดสอบนี้ ถูกกำหนดโดยขนาดของเนื้อที่ของ harddisk

ที่ใช้เก็บข้อมูลว่ามีขนาดที่ว่างที่ใช้เก็บเท่าไร และตัวอักษรต้นแบบในการศึกษาครั้งนี้เป็นตัวอักษรต้นแบบที่ใช้สำหรับกับตัวอักษรแบบ EucrosiaUPC ขนาด 18 point ดังนั้นหากต้องการนำไปใช้กับตัวอักษรในลักษณะอื่นที่มีขนาดเท่ากัน ขั้นตอนการแยกตัวอักษร และการจัดเรียงตัวอักษรสามารถที่ยังคงใช้งานได้ แต่ขั้นตอนการรู้จำตัวอักษร จะมีผลทำให้อัตราการรู้จำที่ถูกต้องมีค่าน้อยลง

4.2 ข้อเสนอแนะ

การศึกษาการรู้จำคำไทยในงานวิจัยนี้เป็นการศึกษาการรู้จำตัวอักษรไทยตัวพิมพ์ที่นำมาเรียงต่อกันเป็นกลุ่มคำหรือประโยค ซึ่งสามารถนำไปใช้ในทางด้านเอกสาร เช่น การพิมพ์เอกสารที่เรามีอยู่ เข้าเก็บไว้ในเครื่องคอมพิวเตอร์ เป็นการทำให้เกิดความสะดวกและรวดเร็วในงานด้านนี้ ซึ่งแม้ในงานวิจัยนี้จะถูกจำกัดในด้านขนาดของของตัวอักษร และรูปแบบของตัวอักษร หากได้รับการพัฒนาให้สามารถใช้กับตัวอักษรได้หลายขนาดและหลาย ๆ รูปแบบของตัวอักษรไทย รวมไปถึงตัวอักษรที่เป็นลายมือเขียน ก็จะเป็นการสะดวกมากขึ้น อีกทั้งหากสามารถทำให้ระบบการรู้จำมีความรวดเร็วในการรู้จำ และความถูกต้องแม่นยำที่สูงขึ้นด้วยแล้ว ก็จะมีประโยชน์เป็นอย่างมากและสามารถนำไปใช้งานได้อย่างกว้างขวางมากขึ้นด้วย

สำหรับการเก็บข้อมูลของตัวอักษรต้นแบบที่ใช้ในการแยกตัวอักษรที่ติดกันเพื่อใช้ในการรู้จำนั้น เป็นองค์ประกอบที่สำคัญ ทั้งนี้เนื่องจากการแยกตัวอักษรที่ติดกันนั้น จำเป็นต้องมีตัวอักษรต้นแบบในการช่วยแยกตัวอักษรให้มีความถูกต้องและแม่นยำ ซึ่งถ้าหากสามารถทำให้ตัวอักษรต้นแบบ 1 ตัวสามารถใช้ได้กับตัวอักษรหลาย ๆ ขนาด ก็จะทำให้ตัวอักษรต้นแบบที่เก็บไว้มีขนาดน้อยลง หรือถ้าหากเก็บเฉพาะตัวอักษรต้นแบบที่คาดว่าจะทำให้เกิดตัวอักษรที่ติดกัน ไม่จำเป็นต้องเก็บหมดทุกตัวอักษร ก็จะทำให้ไม่ต้องเสียเวลาในการอ่านข้อมูลของตัวอักษรต้นแบบ และสำหรับตัวอักษรต้นแบบที่ใช้ในการรู้จำก็เช่นเดียวกัน ถ้าหากสามารถเก็บข้อมูลของตัวอักษรต้นแบบที่ใช้ในการรู้จำนั้นให้ตัวอักษรต้นแบบ 1 ตัวสามารถใช้ได้กับตัวอักษรที่ต้องการรู้จำได้หลายขนาด ก็จะช่วยทำให้ปริมาณตัวอักษรต้นแบบมีน้อยลง ทำให้เสียเวลาในการอ่านข้อมูลตัวอักษรต้นแบบน้อยลง เป็นการเพิ่มความเร็วในระบบรู้จำคำไทยด้วย

สำหรับข้อผิดพลาดที่เกิดขึ้นในงานวิจัยนี้ ซึ่งได้แก่การรู้จำตัวอักษรที่ผิดหรือรู้จำไม่ได้ หากทำการแก้ไขในส่วนของการรู้จำให้มีความถูกต้องมากยิ่งขึ้น ก็จะทำให้ผลของการวิจัยมีความถูกต้องมากยิ่งขึ้น เช่นเพิ่มตัวอักษรต้นแบบเข้าไปสำหรับตัวอักษรที่มีการรู้จำผิด เพื่อเป็นการตรวจสอบตัวอักษร ที่รู้จำผิด การนำเอาโครงสร้างทางภาษาศาสตร์มาช่วยในการตรวจสอบ หรือโครงข่ายนิวรอล (Neural Network) มาใช้ในการเรียนรู้รูปแบบตัวอักษรที่หลากหลาย อีกทั้งขั้นตอนในการแยกกลุ่มของข้อมูลของตัวอักษรที่ติดกันให้มีความถูกต้องมากยิ่งขึ้น ซึ่งขั้นตอนนี้ได้ใช้การเปรียบเทียบกับตัวอักษรต้นแบบ

ที่ได้เก็บไว้ ซึ่งหากตัวอักษรที่นำมาทดสอบมีความคล้ายคลึงกับตัวอักษรต้นแบบก็จะได้ ตัวอักษรที่ถูกต้อง แต่ถ้าหากตัวอักษรที่นำมาทดสอบมีความผิดเพี้ยนไปจากตัวอักษรต้นแบบมาก ก็จะทำให้ไม่สามารถแยกได้ว่าเป็นตัวอักษรอะไร ซึ่งหากขั้นตอนในการแยกตัวอักษรที่ติดกันนี้มีความถูกต้องมาก ยิ่งขึ้นก็จะมีข้อผิดพลาดในการรู้จำน้อยลงด้วยเช่นกัน