



1.1 ความเป็นมาของปัญหา

ปัจจุบันคอมพิวเตอร์ได้เข้ามามีบทบาทในประเทศไทยมากขึ้น จนกระทั่งเครื่องคอมพิวเตอร์เกือบจะกลายมาเป็นเครื่องมือเครื่องใช้ประจำวัน ในหน่วยงานต่างๆ มากมาย สาเหตุเนื่องมาจากปริมาณงานในหน่วยงาน รวมทั้งข้อมูลและข่าวสารที่ทวีมากขึ้น ทำให้ต้องใช้เวลาในการปฏิบัติงาน และค่าใช้จ่ายเกี่ยวกับบุคคลากรสูงขึ้น แต่ในขณะเดียวกัน อุปกรณ์คอมพิวเตอร์กลับมีราคาลดต่ำลง อีกทั้งยังมีโปรแกรมสำเร็จรูปที่ถูกพัฒนาขึ้นมาอย่างพร้อมสมบูรณ์ที่จะให้ประมวผลได้ทันทีตามความต้องการของผู้ใช้ นอกจากนี้แล้ว การนำคอมพิวเตอร์เข้ามาใช้งานยังช่วยให้การประมวลผลข้อมูลทำได้อย่างรวดเร็ว การค้นหาหรือแก้ไขเปลี่ยนแปลงข้อมูลเป็นไปโดยสะดวก ทำให้สามารถนำข้อมูลมาใช้ได้ทันต่อเวลา

จากการที่เครื่องคอมพิวเตอร์ ได้เข้ามามีบทบาทในชีวิตประจำวันมากขึ้นนี้ ทำให้พัฒนาการทางคอมพิวเตอร์ มีแนวโน้มที่จะอำนวยความสะดวกให้กับผู้ใช้ ในด้านการติดต่อสื่อสารกับเครื่องคอมพิวเตอร์มากขึ้น ไม่ว่าจะเป็นการป้อนข้อมูลเข้า หรือการนำผลลัพธ์ออกจากเครื่องคอมพิวเตอร์ให้มีความใกล้เคียงกับประสาทสัมผัสของมนุษย์ พัฒนาการนี้รวมไปถึงการพัฒนาให้เครื่องคอมพิวเตอร์ สามารถรู้จำและสร้างเสียงพูดได้ สามารถอ่านและรู้จำตัวอักษรได้

งานวิจัยนี้จะเน้นเฉพาะ การพัฒนาให้เครื่องคอมพิวเตอร์สามารถอ่านและรู้จำตัวพิมพ์อักษรไทยได้ เพื่ออำนวยความสะดวกให้แก่ผู้ใช้ ในกรณีที่ผู้ใช้ต้องการคัดลอกข้อความที่มีอยู่บนสิ่งพิมพ์เก็บไว้ในอุปกรณ์บันทึกข้อมูลทางคอมพิวเตอร์ ก็สามารถทำได้ง่ายโดยไม่จำเป็นต้องป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์ผ่านทางแป้นพิมพ์ ซึ่งนอกจากจะทำให้เกิดความล่าช้าและเสียเวลาแล้ว ในกรณีที่ผู้ใช้ต้องการเก็บรวบรวมข้อมูลจำนวนมาก ก็จะทำให้ต้องเสียแรงงานคน และค่าใช้จ่ายไปเป็นจำนวนหนึ่งอีกด้วย ดังนั้นจะเห็นได้ว่า หากสามารถพัฒนาให้เครื่องคอมพิวเตอร์อ่านและรู้จำตัวพิมพ์อักษรไทยได้ ก็จะช่วยลดขั้นตอนในการป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์ลงได้ และหลังจากที่ข้อมูลถูกเก็บเอาไว้ในอุปกรณ์บันทึกข้อมูลทางคอมพิวเตอร์เรียบร้อยแล้ว ผู้ใช้ก็สามารถเรียกข้อมูลเหล่านั้นมาตรวจสอบหรือแก้ไขเปลี่ยนแปลงได้โดยง่าย

ในปัจจุบันนี้ การวิจัยเพื่อให้เครื่องคอมพิวเตอร์มีความสามารถในการรู้จำตัวพิมพ์อักษรไทยได้ ส่วนใหญ่มักจะเป็นการวิจัยทางด้านการรู้จำตัวพิมพ์อักษรไทยรูปแบบเดี่ยว (recognition of single-font printed Thai characters) ส่วนทางด้านการรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบ (recognition of multiple-font printed Thai characters) นั้นพบว่ายังมีการวิจัยกันยังไม่มากนัก ซึ่งถ้าจะมองถึงการใช้งานจริงในปัจจุบันแล้วจะเห็นได้ว่า รูปแบบของอักษรที่ปรากฏอยู่บนสิ่งตีพิมพ์ต่างๆ นั้นประกอบไปด้วยรูปแบบของอักษรจำนวนมากมาหลายรูปแบบ ดังนั้นในการวิจัยนี้จึงได้เน้นทางด้านการรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบ เพื่อให้ผลงานวิจัยที่ได้ออกมาสามารถนำไปประยุกต์ใช้กับงานทางด้านการรับข้อมูลอักษรเข้าสู่เครื่องคอมพิวเตอร์ที่มีอยู่เป็นจำนวนมากในชีวิตประจำวันได้

เทคนิคที่นำมาใช้ในงานวิจัยนี้ได้แก่ เทคนิคของการวิเคราะห์เส้นแสดงขอบของอักษร โดยจะนำรหัสทิศทางแบบลูกโซ่ของฟรีแมน กับ ความแตกต่างของทิศทางของเส้นแสดงขอบของอักษร มาใช้ในการตัดแบ่งเส้นแสดงขอบของอักษรออกเป็นส่วนโค้งย่อย ได้แก่ ส่วนโค้งเว้า และส่วนโค้งนูน จากนั้นก็จะเป็นการดึงลักษณะสำคัญของส่วนโค้งย่อยแต่ละส่วนโค้งออกมาสำหรับใช้ในการเปรียบเทียบส่วนโค้งและเปรียบเทียบอักษรต่อไป ในขั้นตอนของการเปรียบเทียบนั้น จะนำเอาเทคนิคของการเปรียบเทียบแบบไดนามิกโปรแกรมมิ่ง ซึ่งเป็นเทคนิคที่นับได้ว่ามีประสิทธิภาพเมื่อนำมาใช้ในการรู้จำเสียงพูดและการรู้จำอักษรในระบบออนไลน์ มาประยุกต์ใช้เพื่อหาความสัมพันธ์ที่คล้ายกันมากที่สุดระหว่างอักษรที่ต้องการรู้จำกับอักษรต้นแบบ และเมื่อได้ความสัมพันธ์ที่คล้ายกันมากที่สุดแล้ว ก็จะใช้ความสัมพันธ์นี้เป็นความสัมพันธ์เริ่มต้นสำหรับการเปรียบเทียบเพื่อหาว่าอักษรต้นแบบใดที่คล้ายกับอักษรที่ต้องการรู้จำมากที่สุด ก็จะถือว่าอักษรนั้นเป็นอักษรที่รู้จำได้ ผลที่ได้จากการวิจัยโดยใช้เทคนิคดังกล่าวมีความถูกต้อง 94.7 เปอร์เซ็นต์ ซึ่งนับว่าอยู่ในเกณฑ์ที่ค่อนข้างสูงเมื่อเทียบกับการวิจัยประเภทเดียวกันแล้ว

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อลดขั้นตอนในการป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์
- 1.2.2 เพื่อประหยัดเวลาและค่าใช้จ่ายในการป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์
- 1.2.3 เพื่ออำนวยความสะดวกแก่ผู้ที่ไม่มีความถนัดในด้านการใช้แป้นพิมพ์ สำหรับป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์ โดยเฉพาะในกรณีที่ข้อมูลมีปริมาณมาก

1.3 ขั้นตอนการวิจัย

- 1.3.1 ศึกษาเทคนิคที่ใช้สำหรับการรู้จำอักขระ
- 1.3.2 วิเคราะห์ และ เลือกเทคนิคที่จะนำมาใช้ในการวิจัยนี้
- 1.3.3 พัฒนาโปรแกรมสำหรับการรับอักขระ และ การรู้จำอักขระ
- 1.3.4 เก็บข้อมูลสำหรับเป็นชุดข้อมูลในการเรียนรู้ เพื่อสร้างเป็นพจนานุกรมของอักขระ
- 1.3.5 เก็บข้อมูลเพื่อการทดสอบโปรแกรม
- 1.3.6 ปรับแก้โปรแกรมเพื่อให้มีประสิทธิภาพมากที่สุด
- 1.3.7 สรุปผลการวิจัย และ ข้อเสนอแนะ

1.4 ขอบเขตการวิจัย

- 1.4.1 พัฒนาสำหรับเครื่องไมโครคอมพิวเตอร์
- 1.4.2 ข้อมูลเข้ากำหนดไว้เป็นตัวพิมพ์อักษรภาษาไทยที่มีลักษณะตัวตรงปกติ 3 รูปแบบ
- 1.4.3 ใช้เครื่องกวาดตรวจด้วยแสง เป็นอุปกรณ์รับข้อมูล
- 1.4.4 รับข้อมูลที่มีรูปแบบ เฉพาะที่เก็บไว้เท่านั้น
- 1.4.5 พัฒนาสำหรับการรู้จำ ในช่วงตั้งแต่มีการตัดอักขระออกจากประโยคทีละตัวแล้ว โดยอักขระที่ตัดออกมาได้นี้มาจากโปรแกรมการตัดอักขระโดยอัตโนมัติ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 คอมพิวเตอร์มีความสามารถในการอ่านตัวพิมพ์อักษรไทยได้
- 1.5.2 การป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์เป็นไปได้อย่างสะดวก
- 1.5.3 ประหยัดเวลาและแรงงานในการป้อนข้อมูลเข้าเครื่องคอมพิวเตอร์