



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โลกเราประกอบด้วยประเทศต่าง ๆ มากกว่า 200 ประเทศ ซึ่งเป็นที่ทราบกันดีว่าทุกประเทศนั้นไม่ได้ใช้ภาษาพูดและภาษาเขียนเดียวกัน ประเทศอังกฤษจะใช้ภาษาอังกฤษและตัวอักษรภาษาอังกฤษ ประเทศจีนจะใช้ภาษาจีนและตัวอักษรภาษาจีน ส่วนประเทศไทยนั้นมีภาษาไทยและตัวอักษรไทยเป็นเอกลักษณ์ประจำชาติ ดังนั้นจะเห็นได้ว่าทั้งภาษาพูดและภาษาเขียนเป็นสิ่งที่จำเป็นในการสื่อสารเท่าเทียมกัน

นักวิจัยเริ่มสนใจศึกษาทำความเข้าใจพร้อมกับสร้างระบบในการรู้จำ ได้แก่ระบบการรู้จำแบบรูปเสียงพูดซึ่งเป็นลักษณะการพยายามจดจำให้ได้ซึ่งเสียงพูดของภาษาที่ศึกษา และระบบการรู้จำแบบรูปตัวอักษรซึ่งเป็นลักษณะพยายามจดจำให้ได้ซึ่งลายมือหรือลักษณะของตัวอักษรของภาษาที่ศึกษา

ประกอบกับปัจจุบันเทคโนโลยีต่าง ๆ ได้รุดหน้าไปมาก อุปกรณ์ต่าง ๆ ที่เคยมีขนาดใหญ่ ในปัจจุบันจะมีขนาดเล็กกระทัดรัดลง เมื่ออุปกรณ์เล็กลงการรับข้อมูลเข้าคอมพิวเตอร์อย่างเช่นแป้นพิมพ์จะมีขนาดเล็กลงและไม่สะดวกในการใช้งาน เมื่อถึงจุดนั้นอุปกรณ์ที่จะนำมาแทนที่แป้นพิมพ์จะกลายมาเป็นปากกาหรือไมโครโฟนแทน หากอุปกรณ์ที่รับข้อมูลเข้าเป็นปากกาหรือระบบที่สามารถรู้จำแบบรูปลายมือเขียนจึงเป็นสิ่งที่ต้องเข้ามาสนับสนุน

การรู้จำแบบรูปตัวอักษรสามารถแบ่งเป็น 2 ชนิดใหญ่ๆ ได้แก่ที่เป็นตัวพิมพ์ และที่เป็นลายมือเขียน ซึ่งเป็นที่ทราบกันดีว่าลักษณะตัวอักษรที่เป็นตัวพิมพ์นั้นเป็นการกำหนดรูปร่างที่แน่นอนคงตัวสำหรับชนิดรูปแบบตัวพิมพ์นั้น ๆ ซึ่งแตกต่างกับลายมือเขียนโดยสิ้นเชิงสิ่งแรกที่เห็นได้ชัดคือการคงตัวของตัวอักษรไม่สามารถกำหนดได้ในตัวอักษรลายมือเขียน แม้กระทั่งคนคนเดียวเขียนตัวอักษรตัวเดียวกันก็ไม่ได้หมายความว่าเหมือนกันทีเดียว อีกทั้งลักษณะตัวอักษรที่เขียนและอุปนิสัยในการเขียนของแต่ละบุคคลยังแตกต่างกัน ด้วยเหตุดังกล่าวจะเห็นได้ว่าการรู้จำแบบรูปตัวอักษรลายมือเขียนจะยากกว่าตัวอักษรตัวพิมพ์และเป็นสิ่งที่ท้าทายในการวิจัยคิดค้นหาวิธีการรู้จำที่มีประสิทธิภาพเข้ามาสนับสนุน

การรู้จำแบบรูปตัวอักษรลายมือเขียนได้มีการศึกษาวิจัยกันมายาวนานกว่า 20 ปี (Ali, F. และ Pavlidis, T. , 1977) โดยเริ่มต้นทำการศึกษาเกี่ยวกับตัวเลขอารบิกก่อนโดยมีจุดประสงค์เพื่อที่จะนำเข้ามาใช้ในการช่วยลดงานในการคัดเลือกจดหมายในกิจการไปรษณีย์ หลังจากที่มีการพัฒนาทางด้านการรู้จำแบบรูปตัวเลขอารบิกประสบความสำเร็จจนได้อัตราความถูกต้องในการรู้จำเป็นที่น่าพอใจแล้วก็เริ่มขยายการศึกษาลงไปในกลุ่ม

ของตัวอักษร ในต่างประเทศนั้นได้ศึกษากันอย่างจริงจังและต่อเนื่องอีกทั้งยังได้รับการสนับสนุนเป็นอย่างดีจึงทำให้การพัฒนาทางด้านความรู้จำแบบรูปลายมือเขียนภาษาอังกฤษนั้นไปไกลถึงความถูกต้องสูงจนสามารถใช้งานได้และเป็นที่ยอมรับ อุปกรณ์ที่อาศัยความรู้จำแบบรูปทางด้านลายมือเขียนนั้นที่สามารถพบได้ในปัจจุบันได้แก่ ปาล์มไฟลิต ซึ่งเป็นอุปกรณ์จดบันทึกสิ่งงานที่ใช้ปากกาในการรับข้อมูลที่เป็นภาษาอังกฤษเข้าไปแล้วทำการแปลงเป็นตัวอักษร ในประเทศไทยนั้นการศึกษาทางด้านลายมือเขียนได้เริ่มมีการวิจัยมาก่อนหน้านี้ (Suraphun Airphaiboon และคณะ, 1989 เป็นต้น) หากแต่ว่าการวิจัยเท่าที่ผ่านมายังขาดความต่อเนื่องและการสนับสนุนเท่าที่ควร หากเราต้องการที่จะมีเครื่องมือที่สามารถรู้จำตัวอักษรภาษาไทย เราจะต้องศึกษาและพัฒนากรรมวิธีในการรู้จำแบบรูปเพื่อให้เหมาะกับภาษาไทย เนื่องจากตัวอักษรภาษาไทยมีลักษณะเฉพาะที่แตกต่างจากภาษาอื่น ๆ

ปัญหาหลักในการรู้จำแบบรูปลายมือเขียน

1. ความไม่คงที่ของตัวอักษร
2. รูปแบบในการเขียน
3. อุณหภูมิในการเขียน
4. และอื่นๆ

ตามที่ได้กล่าวมาเป็นความยากในการรู้จำแบบรูปลายมือเขียน หากเรามีกรรมวิธีที่สามารถแบ่งกลุ่มของตัวอักษรขั้นต้นหรือสามารถดึงตัวอักษรจากกลุ่มที่มีลักษณะเด่นเป็นพิเศษออกมาได้ จะเป็นการช่วยในการทำการจัดกลุ่มได้เป็นอย่างมาก ในกระบวนการรู้จำนั้นหากเราสามารถที่จะทำการจำแนกเบื้องต้นได้ดีแล้ว จะเป็นประโยชน์กับเราในการช่วยลดปริมาณงานลงไปได้มาก อีกทั้งความยุ่งยากและซับซ้อนก็จะลดลง

1.2 วัตถุประสงค์

1. สามารถนำแนววิธีการพิจารณาโดยใช้ลักษณะบ่งความต่าง ช่วยในการจำแนกตัวอักษรพยัญชนะไทย
3. ศึกษาและหาคุณลักษณะบ่งความต่างที่เหมาะสมกับตัวอักษรลายมือเขียนพยัญชนะไทยตัวบรรจง
4. พัฒนากกรรมวิธีการรู้จำแบบรูปสำหรับตัวอักษรพยัญชนะไทยชนิดแยกเดี่ยว

1.3 เป้าหมายและขอบเขตของวิทยานิพนธ์

1. สร้างกรรมวิธีการรู้จำแบบรูปลายมือเขียนอักษรพยัญชนะไทย ก ถึง ฮ จำนวน 44 ตัวอักษร ที่เขียนแบบคัดลายมือลักษณะแยกเดี่ยวตัวตรงปกติได้อัตราการรู้จำไม่ต่ำกว่า 85 เปอร์เซ็นต์
2. พยัญชนะไทยที่ใช้ในการรู้จำเป็นตัวอักษร ก ถึง ฮ จำนวน 44 ตัว ที่เขียนแบบคัดลายมือตามอักขระวิธปกติ ให้เป็นตัวตรงแยกเดี่ยวอยู่ในกรอบที่กำหนด

3. ได้โปรแกรมที่สามารถรู้จำตัวอักษรพยัญชนะไทยชนิดแยกเดี่ยว ซึ่งมีลักษณะดังนี้
 - ทำงานบนระบบปฏิบัติการวินโดวส์ 95 และเขียนขึ้นโดย Ms. Visual C++ 6
 - สามารถทำการเก็บข้อมูลชนิด Portable GrayMap (.pgm)
 - สามารถแสดงผลการรู้จำแบบรูปให้เห็นได้

1.4 ขั้นตอนและวิธีดำเนินงาน

1. ศึกษากรรมวิธีและผลงานที่ผ่านมาในการรู้จำแบบรูปอักษรลายมือเขียน
2. ค้นคว้าและเก็บรวบรวมรายละเอียดเกี่ยวกับตัวอักษรลายมือเขียน
3. ทำการทดสอบแนวความคิดที่น่าเสนอ
4. วิเคราะห์และปรับปรุงผล
5. สร้างวิธีการรู้จำแบบรูปตัวอักษรพยัญชนะไทย
6. สรุปและรวบรวมผลงานวิจัยทั้งหมด พร้อมทั้งจัดทำวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ทราบถึงกระบวนการในการรู้จำแบบรูปลายมือเขียน รวมถึงแนวทางที่เหมาะสมในการพัฒนาเพื่อเพิ่มประสิทธิภาพในการรู้จำแบบรูปลายมือเขียนภาษาไทย
2. เพื่อเพิ่มความสามารถทางด้านความแม่นยำของวิธีการรู้จำแบบรูปลายมือเขียนภาษาไทย
3. สามารถวางแนวทางที่เหมาะสมในการพัฒนาระบบการรู้จำแบบรูปลายมือเขียนภาษาไทยต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย