



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

หากเปรียบเทียบส่วนรับรู้ข้อมูลระหว่างเครื่องคอมพิวเตอร์และมนุษย์แล้ว ส่วนรับรู้ข้อมูลของมนุษย์จะมีความสามารถที่สูงกว่ามากไม่ว่าจะเป็นการรับรู้จากการเห็นภาพ การได้ยิน การสัมผัส การได้กลิ่น หรือการรู้รส เพราะฉะนั้นจึงมีผู้พยายามประดิษฐ์ คิดค้น และพัฒนา ส่วนการรับรู้ของเครื่องคอมพิวเตอร์ให้ทัดเทียมกับมนุษย์ไม่ว่าจะเป็นการพัฒนาให้เครื่องคอมพิวเตอร์สามารถตรวจรู้ข้อมูลทางภาพ (Pattern Recognition) หรือทางเสียง (Sound Recognition) เลียนแบบความสามารถของมนุษย์

การตรวจรู้อักขระ (Character Recognition) เป็นสาขาย่อยหนึ่งของการตรวจรู้ภาพ การนำเทคนิคการตรวจรู้อักขระมาประยุกต์ใช้กับเครื่องคอมพิวเตอร์นับเป็นการพัฒนาให้เครื่องคอมพิวเตอร์สามารถเข้าใจภาษาเขียนที่มนุษย์ใช้ติดต่อสื่อสารกันได้ ซึ่งจะมีผลให้การประมวลผลด้วยเครื่องคอมพิวเตอร์มีความสะดวก รวดเร็ว ถูกต้อง และประหยัดค่าใช้จ่ายกว่าที่เป็นอยู่ในปัจจุบันนี้ เทคนิคการตรวจรู้อักขระด้วยเครื่องคอมพิวเตอร์นี้ยังสามารถนำไปประยุกต์ใช้กับงานวิจัยอื่นๆ ได้มากมายเช่นนำไปใช้กับงานวิจัยด้านการแปลภาษาด้วยเครื่องคอมพิวเตอร์ (Machine Translation) การสร้างอุปกรณ์สำหรับช่วยคนพิการตาบอด การตรวจสอบความถูกต้องของเอกสารทางธุรกิจ การแยกพัสดุไปรษณีย์ เป็นต้น

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อให้เครื่องคอมพิวเตอร์สามารถตรวจรู้อักขระภาษาไทยแต่ละรูปได้

1.2.2 ำให้เครื่องคอมพิวเตอร์แทนค่าอักขระภาษาไทยแต่ละรูปด้วยรหัสซึ่งสามารถนำไปประมวลผลได้

1.2.3 เพื่อเป็นแนวทางในการพัฒนาการตรวจรู้อักขระภาษาไทยด้วยเครื่องคอมพิวเตอร์ให้สมบูรณ์ยิ่งขึ้น

1.3 ขอบเขตของงานวิจัย

เนื่องจากการพัฒนาเทคโนโลยีเรื่องการตรวจรู้อักขระนี้จะต้องอาศัยเทคโนโลยีทั้งทางด้านแสง (Optics) อิเล็กทรอนิกส์ (Electronics) เครื่องกล (Mechanics) และคอมพิวเตอร์ (Computer) ประกอบกัน แต่ในการวิจัยนี้จะมีมุ่งพัฒนาเฉพาะเทคโนโลยีทางด้านคอมพิวเตอร์เรื่องการตรวจรู้อักขระเท่านั้น โดยแบ่งการวิจัยออกเป็น 3 ส่วนคือ ส่วนแปลงอักขระที่ต้องการตรวจรู้ให้เหลือเพียงโครงร่าง ส่วนแปลงโครงร่างของอักขระให้เป็นรหัสที่จะนำไปใช้ตรวจรู้อักขระและ ส่วนค้นหา (Search) ผลลัพธ์ด้วยการเปรียบเทียบรหัสจากส่วนที่สองกับรหัสต้นแบบ รหัสต้นแบบเป็นรหัสที่สร้างไว้ใช้ค้นหาผลลัพธ์ของการตรวจรู้

อักขระนั้นว่า

```

00000000000000000000
00000000000000000000
00000000000000000000
00001111111110000000
00111111111111000000
01111000000111100000
00011100000011100000
00111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
0111000000011100000
00000000000000000000
00000000000000000000
00000000000000000000

```

รูปที่ 1.1

แสดงภาพบิตเมทริกซ์ตัวอักษร "ก"

อักขระที่ใช้ในการประมวลผลนี้จะใช้อักขระที่เป็นภาพบิตเมทริกซ์ตัวอักษร (Bit Image Matrix) (รูปที่ 1.1) ซึ่งเป็นอักขระที่สร้างเลียนแบบผลลัพธ์ของเครื่องแปลงสัญญาณภาพให้เป็นดิจิทัล (Digital) รายละเอียดของภาพบิตเมทริกซ์ตัวอักษรมีดังนี้

1.3.1 ใช้ตัวอักขระซึ่งมีรูปแบบ (Pattern) เป็นตัวพิมพ์ตามที่ใช้ในเครื่องพิมพ์ดีดทั่วไปในการตรวจรู้

1.3.2 ใช้ตัวอักขระที่อยู่ในรูปเมทริกซ์ขนาด 20 X 20 หน่วย

1.3.3 ลักษณะของตัวอักขระจะต้องเป็นแบบตัวพิมพ์ตั้งตรงไม่หัวกลับหรือตะแคง

1.3.4 อักขระที่จะตรวจรู้มีจำนวนทั้งหมด 70 รูป โดยแบ่งเป็นพยัญชนะ จำนวน 44 รูป สระ จำนวน 21 รูป วรรณยุกต์อีกจำนวน 4 รูป และอักษรว่าง (Blank) อีก 1 รูป (รูปที่ 1.2)

1.3.5 จำนวนแบบพิมพ์ (Fonts) ของอักขระต้นแบบที่จะใช้ในการตรวจรู้จะใช้อักขระต้นแบบภาษาไทยจำนวน 5 แบบพิมพ์ อักขระต้นแบบภาษาอังกฤษจำนวน 1 แบบพิมพ์ (รูปที่ 1.3 - 1.8) และอักขระทดสอบภาษาไทยอีกจำนวน 2 แบบพิมพ์ (ภาคผนวก ค.)

ก ข ช ค ศ ฆ ง จ ฉ ช ช ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ศ ษ ส ห ฬ อ ฮ พยัญชนะ
ะ ำ ั ิ ึ ุ ู ใ ใ โ ใ ำ ฤ ฌ ั ิ ุ ู ใ ใ สระ
ฯ ๗ ๘ ๙ วรรณยุกต์

รูปที่ 1.2 แสดงอักขระภาษาไทย



รูปที่ 1.3 แสดงอักษรต้นแบบภาษาไทยแบบพิมพ์ที่ 1



รูปที่ 1.4 แสดงอักขระต้นแบบภาษาไทยแบบพิมพ์ที่ 2

ก	ข	ฃ	ค	ฅ	ฉ	ช	จ
ฉ	ช	ฃ	ฅ	ง	ฉ	ซ	ฌ
ฌ	ญ	ฎ	ฏ	ฐ	ฑ	ฒ	ณ
ณ	ด	ต	ถ	ท	ธ	น	บ
ป	ผ	ฝ	พ	ฟ	ภ	ม	ย
ร	ฤ	ล	ฦ	ว	ศ	ษ	ส
ห	ฬ	อ	ฮ	ฯ	ะ	ั	า
ำ	ิ	ี	ึ	ื	ุ	ู	ฺ
฻	฼	฾	฿	๐	๑	๒	๓
๔	๕	๖	๗	๘	๙	๐	๑

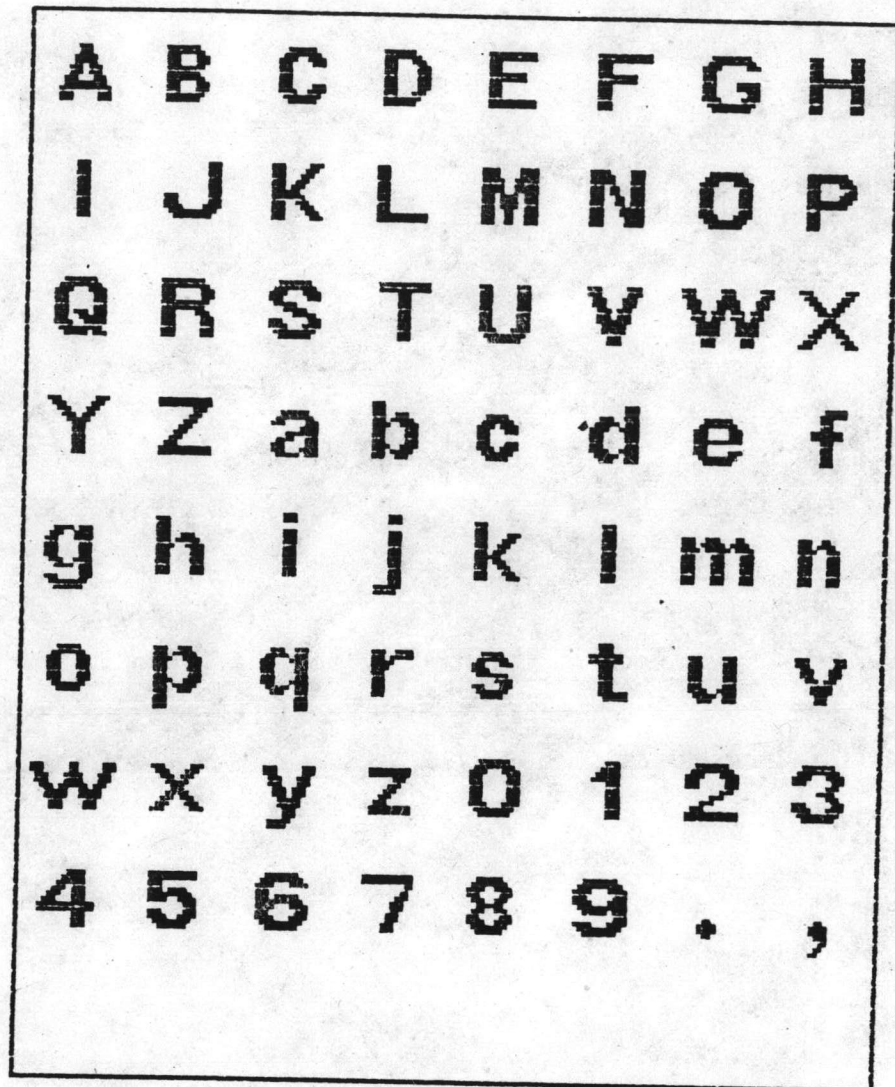
รูปที่ 1.5 แสดงอักขระต้นแบบภาษาไทยแบบพิมพ์ที่ 3



รูปที่ 1.6 แสดงอักขระต้นแบบภาษาไทยแบบพิมพ์ที่ 4



รูปที่ 1.7 แสดงอักขระต้นแบบภาษาไทยแบบพิมพ์ที่ 5



รูปที่ 1.8 แสดงอักขระต้นแบบภาษาอังกฤษแบบทิมพ์ที่ 1

ผลลัพธ์ที่ได้จากการตรวจรู้อักขระภาษาไทยนี้จะเก็บเป็นรหัสแอสกี (ASCII Code) โดยจะเก็บรหัสนั้นเป็นแถว (String) ต่อกันในระดับเดียวกัน

อุปกรณ์คอมพิวเตอร์ที่ใช้ในการวิจัยผลนี้จะใช้เครื่องไมโครคอมพิวเตอร์ 16 บิต (bits) ขนาด 512 KB มีดิสเกตต์ไดรฟ์ (Diskette Drive) ขนาด 5 1/4 นิ้ว 2 คู่ จานแม่เหล็ก (Harddisk) ขนาด 10 MB และเครื่องพิมพ์ (Printer) 1 เครื่อง

1.4 วิธีดำเนินการวิจัย

1.4.1 ศึกษาถึงวิธีการที่จะแยกความแตกต่างของอักขระภาษาไทยแต่ละรูป

1.4.2 วิเคราะห์ถึงวิธีการที่จะใช้เครื่องคอมพิวเตอร์ตรวจรู้อักขระภาษาไทย โดยใช้วิธีการแยกความแตกต่างของอักขระภาษาไทย จากข้อ 1.4.1

1.4.3 สร้างและพัฒนาชุดคำสั่งสำหรับตรวจรู้อักขระภาษาไทยพร้อมทั้งแทนค่าอักขระที่ตรวจรู้ได้นั้นด้วยรหัสซึ่งสามารถนำไปใช้ประมวลผลได้

1.4.4 ทดสอบการทำงานของชุดคำสั่งที่พัฒนาขึ้น

1.4.5 สรุปผลการวิจัย