

ผลการวิจัย

การวิจัยในหัวข้อเรื่อง ระบบออนไลน์สำหรับการรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบนี้ ได้นำรูปแบบของอักขระจำนวน 3 รูปแบบ มาเป็นตัวอย่างสำหรับการวิจัย โดยแต่ละรูปแบบของอักขระมีขนาด 40 x 40 จุด ซึ่งเป็นขนาดของตัวพิมพ์ที่ใช้กันทั่วไปในงานพิมพ์ต่างๆ

4.1 ตัวอย่างอักขระที่ใช้ในการวิจัย

รูปแบบอักขระที่ใช้ในการวิจัยได้มาจากรูปแบบอักขระซึ่งมีอยู่ในโปรแกรม Pagemaker โดยแต่ละรูปแบบมีขนาดเป็น 40 x 40 จุด แต่ละรูปแบบของอักขระดังแสดงในรูปที่ 4.1(ก) รูปที่ 4.1(ข) และ รูปที่ 4.1(ค) จะถูกพิมพ์ออกมา โดยผ่านเครื่องพิมพ์ชนิดเลเซอร์ ทำให้ได้ตัวพิมพ์ที่มีลักษณะของเส้นคมชัดซึ่งจะช่วยให้ความถูกต้องในการรู้จำสูงขึ้น

ก ข ค ง จ ฉ ช ซ

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 พ อ ฮ

ะ ำ

ั ึ ู ุ

เ ื ุ

เ แ โ ใ ไ ำ ุ

ฤ ี ุ ู ุ + ุ ุ

๑ ๑ ๑

รูปที่ 4.1(ก) อักขระรูปแบบที่ 1

ก ข ค ฅ ง จ ฉ ช ซ
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 พ อ ฮ

ะ ำ

ก ข ค ฅ ง จ ฉ ช ซ

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 พ อ ฮ

รูปที่ 4.1(ข) อักษรรูปแบบที่ 2

ก ข ค ฅ ง จ ฉ ช ซ

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 พ อ ฮ

ะ ำ

ก ข

ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ

ด ต ถ ท ธ น บ ป

พ ฟ ภ

รูปที่ 4.1(ค) อักษรรูปแบบที่ 3

4.2 ผลการวิจัย

4.2.1 อุปกรณ์ที่ใช้ในการวิจัย

4.2.1.1 เครื่องพิมพ์ชนิดเลเซอร์

สำหรับใช้ในการพิมพ์รูปแบบข้อมูลของอักขระทั้งสามรูปแบบ เพื่อให้ลักษณะของตัวพิมพ์ที่ได้มีความคมชัดของเส้น ซึ่งจะมีส่วนช่วยทำให้ความถูกต้องในการรู้จำสูงขึ้น

4.2.1.2 เครื่องกวาดตรวจด้วยแสง

เป็นอุปกรณ์ที่ใช้ในการรับข้อมูลเข้าสู่เครื่องคอมพิวเตอร์ โดยสามารถรับข้อมูลเข้ามาได้ทีละ 1 หน้ากระดาษ

4.2.1.3 เครื่องไมโครคอมพิวเตอร์

เป็นอุปกรณ์ที่ใช้ควบคุม และ จัดการเกี่ยวกับการทำงานของทั้ง เครื่องพิมพ์ชนิดเลเซอร์และเครื่องกวาดตรวจด้วยแสง ข้อมูลที่รับเข้ามาจากเครื่องกวาดตรวจด้วยแสงก็จะถูกบันทึกไว้ในหน่วยความจำของเครื่องไมโครคอมพิวเตอร์นี้ และยังใช้เป็นอุปกรณ์สำหรับการประมวลผลเพื่อการรู้จำอักขระอีกด้วย

4.2.2 จำนวนอักขระที่ใช้ในการวิจัย

โดยทั่วไปอักขระไทยที่ใช้เป็นพื้นฐานเพื่อการรู้จำอักขระจะมีจำนวน 67 ตัว ดังที่ได้กล่าวไว้แล้วในหัวข้ออักขระไทยในบทที่ 2 ในการวิจัยนี้ได้ใช้อักขระจำนวน 3 รูปแบบ โดยอักขระแต่ละตัวของแต่ละรูปแบบจะถูกกวาดตรวจด้วยแสงจำนวน 5 ครั้ง

$$\begin{aligned} & \text{ดังนั้นอักขระที่ใช้ในการวิจัยทั้งหมดควรมีจำนวน} \\ & = 3 \text{ รูปแบบ} \times (5 \text{ ครั้ง} \times 67 \text{ ตัว}) \\ & = 1005 \text{ ตัว} \end{aligned}$$

แต่เนื่องจากมีอักษรจำนวน 2 ตัว คือ สระ แ และ สระ ำ เมื่อนำ มากวาดตรวจด้วยแสงแล้ว ปรากฏว่า อักษรรูปแบบที่ 1 และ อักษรรูปแบบที่ 3 จะได้ สระ แ แยกออกมาเป็น สระ เ 2 ตัว เนื่องจากตัวพิมพ์สระ แ ของทั้งสองรูปแบบนี้มีลักษณะที่ เหมือนกับเป็น สระ เ 2 ตัว ที่อยู่ชิดกันมาก (แต่ไม่ได้ยึดติดกัน) ดังนั้นในขั้นตอนของการตัด อักษรโดยอัตโนมัติ นั้น สระ แ จึงถูกตัดแบ่งออกมาเป็น สระ เ 2 ตัว เนื่องจากอัลกอริทึม ในการตัดแบ่งอักษรจะถือว่า ถ้ามีช่องว่างระหว่างอักษรแม้เพียง 1 ช่อง ก็จะถือว่าเป็นอักษร อีกตัวหนึ่ง ส่วนสระ ำ ก็จะถูกแยกออกมาเป็น 2 ตัว คือเป็น สระ ิ กับ สระ ำ ด้วยเหตุผล เช่นเดียวกับกับการแยกกันของ สระ แ ดังนั้นทั้งสองรูปแบบนี้จึงประกอบไปด้วยอักษรทั้งหมด 69 ตัว สำหรับอักษรรูปแบบที่ 2 นั้น สระ ำ ก็จะถูกแยกออกมาเป็น 2 ตัว เช่นกัน คือ เป็น สระ ิ กับ สระ ำ ดังนั้นรูปแบบนี้จึงประกอบไปด้วยอักษรทั้งหมด 68 ตัว

สรุปได้ว่า อักษรทั้งหมดสำหรับอักษรรูปแบบที่ 1 มีจำนวน = 69 ตัว
 อักษรทั้งหมดสำหรับอักษรรูปแบบที่ 2 มีจำนวน = 68 ตัว
 อักษรทั้งหมดสำหรับอักษรรูปแบบที่ 3 มีจำนวน = 69 ตัว

ด้วยเหตุนี้อักษรที่ใช้ในการวิจัยทั้งหมดจึงมีจำนวน

$$= [2 \times (69 \times 5)] + (68 \times 5)$$

$$= 1030 \text{ ตัว}$$

4.2.3 จำนวนอักษรในพจนานุกรมของอักษร

เนื่องจากแต่ละรูปแบบของอักษรจะถูกกวาดตรวจด้วยแสงจำนวน 5 ครั้ง ดังนั้นจึงได้นำรูปแบบของอักษรจำนวน 1 ครั้ง ใน 5 ครั้ง ของการกวาดตรวจด้วยแสง มาสร้างเป็นพจนานุกรมของอักษร นั่นก็คือในพจนานุกรมของอักษรก็จะประกอบไปด้วยรูปแบบ ของอักษรทั้งสามรูปแบบ เนื่องจากในการวิจัยนี้เป็นการวิจัยเพื่อการศึกษาอักษรหลายรูปแบบ (ในการวิจัยนี้ใช้ 3 รูปแบบ) จึงควรจะได้เก็บทุกรูปแบบของอักษรที่ใช้ในการวิจัยไว้ในพจนานุกรม ของอักษรเพื่อที่ว่าเมื่อต้องการรู้จำอักษรรูปแบบใดรูปแบบหนึ่งในสามรูปแบบนี้ ก็จะทำให้ผลที่ ได้จากการรู้จำมีเปอร์เซ็นต์ของความถูกต้องมากขึ้น กว่าการที่จะเก็บเพียงรูปแบบใดรูปแบบหนึ่ง หรือเพียงสองรูปแบบเท่านั้น

ดังนั้นพจนานุกรมของอักษรจึงประกอบด้วยอักษรทั้งหมด

$$= 69 + 68 + 69$$

$$= 206 \text{ ตัว}$$

จากการวิจัยพบว่า ถ้าให้มีการเรียนรู้อักขระที่จะนำมาใช้ในการรู้จำแล้ว ก็จะทำให้ผลของความถูกต้องในการรู้จำเพิ่มสูงขึ้น เนื่องจากว่าถ้ามีอักขระที่มีลักษณะแตกต่างออกไปจากที่เก็บไว้ในพจนานุกรมของอักขระมากกว่าค่าที่กำหนดไว้ค่าหนึ่งแล้ว อักขระนั้นก็ถูกเก็บเพิ่มเข้าไว้ในพจนานุกรมของอักขระด้วย ทำให้ในพจนานุกรมของอักขระหลังจากที่มีการเรียนรู้อักขระที่จะนำมาใช้ในการรู้จำแล้วมีจำนวนอักขระเพิ่มมากขึ้น

$$\begin{aligned} & \text{ดังนั้นพจนานุกรมของอักขระหลังจากมีการเรียนรู้จะมีจำนวนอักขระเป็น} \\ & = 206 + 25 \\ & = 231 \text{ ตัว} \end{aligned}$$

4.2.4 ผลการรู้จำอักขระ

ในการวิจัยเพื่อการรู้จำตัวพิมพ์อักขระไทยหลายรูปแบบนี้ ได้ใช้อักขระขนาด 40×40 จุด จำนวน 3 รูปแบบ แต่ละรูปแบบนำไปกวาดตรวจด้วยแสงจำนวน 5 ครั้ง คิดเป็นจำนวนอักขระที่ใช้ในการวิจัยทั้งหมด 1030 ตัว และเนื่องจากในการวิจัยใช้รูปแบบอักขระจำนวน 3 รูปแบบ ดังนั้นในพจนานุกรมของอักขระจึงได้เก็บรูปแบบอักขระทั้งสามรูปแบบนั้นไว้ สำหรับใช้เป็นตัวแทนของอักขระแต่ละรูปแบบ โดยแต่ละรูปแบบของอักขระจะได้มาจากการกวาดตรวจด้วยแสงจำนวน 1 ครั้ง จากการกวาดตรวจทั้งหมด 5 ครั้ง คิดเป็นจำนวนอักขระในพจนานุกรมของอักขระทั้งหมดหลังจากการเรียนรู้แล้ว 231 ตัว

เนื่องจากในการวิจัยนี้ได้นำวิธีการแบบไดนามิกโปรแกรมมิง มาใช้ในขั้นตอนของการเปรียบเทียบเพื่อการรู้จำอักขระ ดังนั้นในขั้นแรกของการวิจัยจึงต้องมีการทดสอบเพื่อเปรียบเทียบประสิทธิภาพระหว่างสมการแบบไดนามิกโปรแกรมมิงแบบสมมาตร และ แบบไม่สมมาตร เสียก่อน เพื่อที่จะหาว่าสมการแบบใดและค่าความชันเท่าใดในตารางที่ 2.1 ที่จะเหมาะสมที่สุดสำหรับนำมาใช้ในการวิจัยนี้ โดยอักขระที่ใช้ในการทดสอบนี้ได้มาจากการนำพยัญชนะไทย 42 ตัว มากวาดตรวจด้วยแสงจำนวน 5 ครั้ง คิดเป็นจำนวนอักขระทั้งหมดที่ใช้ในการทดสอบ 210 ตัว ผลที่ได้จากการทดสอบแสดงได้ดังตารางที่ 4.1 ซึ่งเมื่อพิจารณาผลการทดสอบที่ได้นี้แล้วจะเห็นได้ว่า สมการไดนามิกโปรแกรมมิงแบบไม่สมมาตรโดยใช้ค่าความชันเป็น 1 ($P = 1$) เป็นสมการที่เหมาะสมที่สุดสำหรับนำมาใช้ในการวิจัยนี้ต่อไป เนื่องจากมีจำนวนอักขระที่ผิดพลาดจากการรู้จำเป็นจำนวนน้อยที่สุดเมื่อเทียบกับสมการแบบอื่นๆ แล้ว

ตารางที่ 4.1 ผลการทดสอบเพื่อเปรียบเทียบประสิทธิภาพระหว่างสมการไดนามิกโปรแกรมมิ่งแบบสมมาตรและแบบไม่สมมาตร โดยใช้ค่าความชัน (P) ที่แตกต่างกัน

จำนวน อักขระ (ตัว)	P							
	0		1/2		1		2	
	สมมาตร	ไม่สมมาตร	สมมาตร	ไม่สมมาตร	สมมาตร	ไม่สมมาตร	สมมาตร	ไม่สมมาตร
ที่รู้จักได้	201	197	199	204	198	205	193	201
ผิดพลาด	9	13	11	6	12	5	17	9

ในการทําวิจัยนี้ เมื่อสามารถหาลักษณะสำคัญของอักขระที่ต้องการรู้จักได้แล้ว ก็จะนำลักษณะสำคัญเหล่านั้นมาเปรียบเทียบกับอักขระต้นแบบในพจนานุกรมของอักขระ โดยการเปรียบเทียบในขั้นแรกจะเป็นการเปรียบเทียบแบบหยาบก่อน คือ เป็นการหาอักขระต้นแบบที่มีค่าอัตราส่วนความกว้างต่อความสูงของอักขระแตกต่างจากค่าอัตราส่วนความกว้างต่อความสูงของอักขระที่ต้องการรู้จักไม่เกินค่าที่กำหนดไว้ค่าหนึ่ง โดยที่อักขระต้นแบบนี้จะต้องมีจำนวนหัวและจำนวนส่วนโค้งเท่ากันกับอักขระที่ต้องการรู้จักด้วย จากนั้นก็จะหาค่าความแตกต่างระหว่างทุกคู่ส่วนโค้งของอักขระทั้งสองนั้น เพื่อที่จะหาให้ได้ว่าส่วนโค้งคู่ใดเป็นส่วนโค้งคู่ที่มีความคล้ายกันมากที่สุด เมื่อได้คู่ส่วนโค้งที่คล้ายกันมากที่สุดระหว่างอักขระที่ต้องการรู้จักกับอักขระต้นแบบแล้ว ก็จะใช้คู่ส่วนโค้งนี้เป็นคู่ส่วนโค้งเริ่มต้น สำหรับใช้ในการหาค่าความแตกต่างระหว่างอักขระที่ต้องการรู้จักกับอักขระต้นแบบต่อไป

จากการวิจัยที่ได้ทํามาทั้งหมดโดยใช้เทคนิคที่ได้กล่าวไปข้างต้นแล้วนั้นสามารถสรุปผลการรู้จักตัวพิมพ์อักขระไทยหลายรูปแบบได้ ดังแสดงในตารางที่ 4.2 และ ตารางที่ 4.3

ตารางที่ 4.2 ผลการรู้จำตัวพิมพ์อักษรไทยในแต่ละรูปแบบ (จำนวนอักษร)

รูปแบบ ของ อักษร	จำนวนอักษร (ตัว)			
	ทั้งหมด	รู้จำได้	ผิดพลาด	ไม่รู้จำ
รูปแบบที่ 1	345	329	15	1
รูปแบบที่ 2	340	316	21	3
รูปแบบที่ 3	345	330	14	1
รวม	1030	975	50	5

จากผลการวิจัยในตารางที่ 4.2 สามารถสรุปผลการรู้จำของตัวพิมพ์อักษรไทยหลายรูปแบบ (ในรูปของเปอร์เซ็นต์) ได้ดังในตารางที่ 4.3

ตารางที่ 4.3 ผลการรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบ (เปอร์เซ็นต์)

ลักษณะการรู้จำของอักษร	จำนวนอักษร(ตัว)	เปอร์เซ็นต์
สามารถรู้จำได้ (ถูกต้อง)	975	94.7 %
ไม่สามารถรู้จำได้ (ผิดพลาด)	50	4.8 %
ไม่สามารถบอกได้ว่าเป็นอักษรใด (ไม่รู้จำ)	5	0.5 %
จำนวนอักษรที่ใช้ในการรู้จำทั้งหมด	1030	100 %

จากผลการวิจัยในตารางที่ 4.2 และ ตารางที่ 4.3 สามารถสรุปได้ว่า ใน การวิจัยเพื่อการรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบซึ่งใช้อักษรจำนวน 3 รูปแบบ โดยอักษรที่

ใช้มีขนาด 40 x 40 จุด แต่ละรูปแบบนำไปกวาดตรวจด้วยแสงจำนวน 5 ครั้ง คิดเป็นจำนวนอักษรที่ใช้ในการวิจัยทั้งหมด 1030 ตัว และมีจำนวนอักษรในพจนานุกรมของอักษรหลังจากการเรียนรู้แล้ว 231 ตัว เมื่อนำอักษรทั้งหมดไปผ่านขั้นตอนการของความรู้จำอักษร ซึ่งใช้เทคนิคของการวิเคราะห์เส้นแสดงขอบของอักษร ในการดึงลักษณะสำคัญของอักษรออกมา และใช้วิธีการของไดนามิกโปรแกรมมิ่ง ในการเปรียบเทียบอักษรที่ต้องการรู้จำกับอักษรต้นแบบ ผลที่ได้จากการวิจัยมีความถูกต้อง 94.7 เปอร์เซ็นต์



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย