

บทที่ 1

บทนำ

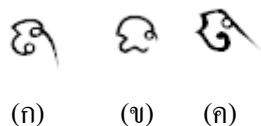
1.1 ความเป็นมาและความสำคัญของปัญหา

มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย วิทยาเขตขอนแก่น ได้เปิดดำเนินการเรียนการสอนมา ตั้งแต่ปีการศึกษา 2529 เป็นต้นมา โดยในช่วงแผนพัฒนาวิทยาเขตขอนแก่นระยะที่ 9 (พ.ศ. 2545-2549) วิทยาเขตขอนแก่นได้กำหนดวิสัยทัศน์ไว้ว่า “ในปีพุทธศักราช 2549 วิทยาเขตขอนแก่นจะเป็นศูนย์กลาง การศึกษาสำหรับพระภิกษุ สามเณร และคฤหัสถ์ ในภาคตะวันออกเฉียงเหนือและประเทศเพื่อนบ้านที่มี ทั้งคุณภาพของบุคลากรและประสิทธิภาพในการให้บริการวิชาการด้านพระพุทธศาสนา” และวิทยาเขต ขอนแก่นก็ได้ดำเนินการพันธกิจของมหาวิทยาลัย ดังนี้

- 1) ผลิตบัณฑิต
- 2) วิจัยและพัฒนา
- 3) ส่งเสริมพระพุทธศาสนาและบริการวิชาการแก่สังคม
- 4) ทำนุบำรุงศิลปวัฒนธรรม

จากรายงานการวิจัย เรื่อง 20 ปี : ภาพลักษณ์วิทยาเขตขอนแก่นในทัศนะของคณะสงฆ์ภาค 9 [1] สรุปว่า กลุ่มตัวอย่างส่วนมากให้คะแนนทัศนะต่อภาพลักษณ์วิทยาเขตขอนแก่นเกี่ยวกับการมีส่วนร่วมในการ ส่งเสริมวัฒนธรรมในท้องถิ่น คิดเป็นร้อยละ 88.1 และวิทยาเขตขอนแก่นสามารถเป็นศูนย์กลางของการ เผยแพร่พระพุทธศาสนาได้ คิดเป็นร้อยละ 85.1 ซึ่งทำให้วิทยาเขตขอนแก่นมีภาพลักษณ์ที่โดดเด่นในด้าน ของการส่งเสริมวัฒนธรรมท้องถิ่นและศูนย์กลางของการเผยแพร่พระพุทธศาสนา

ในการวิจัย เรื่อง การรู้จำอักขรธรรมอีสานโดยใช้ตัวแบบฮิดเดนมาร์คอฟ (E-Sarn Dharma Character Recognition using Hidden Markov models) เป็นการศึกษาวิจัยเกี่ยวกับรูปแบบตัวอักษร(Font) ของอักขระอักขรธรรมอีสานที่มีหน่วยงานต่างๆ ได้คิดค้นและพัฒนาสำหรับการพิมพ์บนเครื่อง คอมพิวเตอร์ ซึ่งทำให้เกิดผลดีในด้านของการอนุรักษ์มรดกและรักษาภูมิปัญญาของบรรพบุรุษให้คงอยู่คู่ แผ่นดินต่อไป สนองต่อพันธกิจของมหาวิทยาลัยและตรงตามภาพลักษณ์ที่สังคมได้ให้มุมมอง นอกจากนี้ แล้วยังทำให้เทคโนโลยีของการพิมพ์หนังสือที่เป็นตัวอักขรธรรมอีสานมีความสะดวกมากยิ่งขึ้นกว่าเดิม ที่เดิมที่จะเป็นลักษณะของการเขียนด้วยลายมือ และได้มีการพิมพ์เป็นหนังสือออกเผยแพร่หลายเล่ม แต่ ปัญหาที่มีอยู่ว่า แต่ละหน่วยงานที่พัฒนาแบบตัวอักขรธรรมนั้น อาจจะมีรูปแบบของอักขระที่แตกต่างกัน ไป ตามท้องถิ่น หรือตามที่ได้รู้เรียนมา ตัวอย่างเช่น ตัว ช ซึ่งมีแบบอักขระที่ไม่เหมือนกัน ดังภาพที่ 1.1



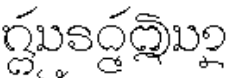
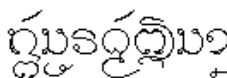
ภาพที่ 1.1 แสดงอักขระอักษรธรรมอีสานตัวอักษร ข โดย

ภาพ (ก) จาก มจร.ขอนแก่น ภาพ (ข) จาก ม.ราชภัฏอุบลฯ และภาพ (ค) จาก ม.มหาสารคาม

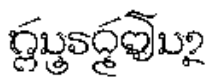
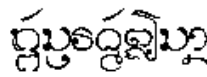
จากปัญหาดังกล่าวข้างต้น จึงได้มีการศึกษาคุณลักษณะสำคัญของอักขระ(Character)อักษรธรรมอีสานโดยใช้กระบวนการรู้จำในระบบคอมพิวเตอร์ โดยการเรียนรู้อักขระธรรมอีสานนั้น เป็นการนำเอาลักษณะของตัวอักขระอักษรธรรมอีสานแต่ละตัวมาใส่ในระบบการเรียนรู้ของคอมพิวเตอร์ เพื่อให้คอมพิวเตอร์รู้ว่าตัวอักษรที่ input เข้าไปนี้ เป็นตัวอะไร ซึ่งจะเป็นประโยชน์ต่อไปในการนำไปจัดทำโปรแกรมแปลงไฟล์ภาพเอกสารไปเป็นไฟล์ข้อความ หรือที่รู้จักกันดีในนามของ OCR (Optical Character Recognition) ระบบการเรียนรู้ของคอมพิวเตอร์เป็นที่รู้จักกันดีในนามของ การรู้จำ (Recognition) ซึ่งเป็นศาสตร์แขนงหนึ่งของสาขาวิทยาการคอมพิวเตอร์ในวิชา ปัญญาประดิษฐ์ (Artificial Intelligence) โดยทั่วไประบบการรู้จำสามารถแบ่งออกได้หลายประเภท เช่น การรู้จำลายนิ้วมือ การรู้จำภาพใบหน้า การรู้จำเสียง เป็นต้น ซึ่งก่อนที่จะเข้าสู่กระบวนการรู้จำนั้นจะต้องผ่านการประมวลผลหลายขั้นตอนเพื่อลดความซับซ้อนของระบบและเพิ่มประสิทธิภาพของการทำงาน เช่น กระบวนการ Noise Remove, Image Enhancement, Edge Detection และ Thinning เป็นต้น [2] ดังนั้นงานวิจัยนี้ จึงได้นำเสนอวิธีการหนึ่งซึ่งสามารถนำไปประยุกต์ใช้งานได้ในระบบการรู้จำอักขระอักษรธรรมอีสาน โดยสามารถทำให้ระบบจดจำภาพอักขระและสามารถระบุได้อย่างชัดเจนว่าภาพอักขระนั้นเป็นอักขระอักษรธรรมอีสานตัวใด ซึ่งจะทำให้สามารถจัดเก็บอักขระนั้นลงในระบบคอมพิวเตอร์ได้โดยอัตโนมัติ ซึ่งในปัจจุบันได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ให้สามารถรู้จำตัวอักขระในภาษาต่างๆ เช่น ภาษาอังกฤษ, ภาษาจีน และอื่นๆ โดยใช้คอมพิวเตอร์ในการอ่านข้อมูลรูปภาพตัวอักขระและแปลงข้อมูลภาพไปเป็นข้อความได้อย่างมีประสิทธิภาพ แต่ยังไม่มีการพัฒนาโปรแกรมคอมพิวเตอร์ให้สามารถรู้จำตัวอักขระอักษรธรรมอีสานได้ ดังนั้นผู้วิจัยจึงได้นำเสนอวิธีการในการรู้จำ โดยได้ทำการเปรียบเทียบประสิทธิภาพของการรู้จำโดยอาศัยการวัดและปรับสัดส่วนของค่าการกระจายความหนาแน่นของจุดเพื่อใช้เป็นลักษณะเฉพาะของตัวอักขระที่ผ่านการปรับมาตรฐานของอักขระแบบไม่เชิงเส้น จากนั้นจึงใช้ตัวแบบฮิดเดนมาร์คอฟเป็นกลไกหลักในกระบวนการรู้จำ

สถานศึกษาต่างๆ 5 แบบ รวมกับที่สร้างขึ้นมาเอง 1 แบบ รวมทั้งสิ้นมีตัวพิมพ์อักษรธรรมอีสาน 6 แบบ แยกเป็นดังนี้

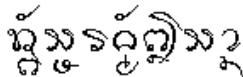
- 1) โรงเรียนสว่างวีระวงศ์ จ.อุบลราชธานี 2 แบบ พัฒนาโดย อาจารย์วัฒน์ ศรีสว่าง

ตัวอย่าง  

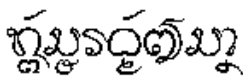
- 2) มหาวิทยาลัยราชภัฏอุบลราชธานี 2 แบบ พัฒนาโดย ผศ.สันติย์ โกคาพันธ์

ตัวอย่าง  

- 3) มหาวิทยาลัยมหาสารคาม 1 แบบ พัฒนาโดย นายสัญญา วุฒิสาร

ตัวอย่าง 

- 4) มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย วิทยาเขตขอนแก่น 1 แบบ พัฒนาโดย
นายนิรันดร์ เลิศวีรพล (ผู้วิจัย)

ตัวอย่าง 

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อออกแบบและสร้างแบบตัวอักษรธรรมอีสานบนคอมพิวเตอร์
- 1.2.2 เพื่อศึกษาขั้นตอนและวิธีการรู้จำตัวอักษรธรรมอีสาน โดยใช้ตัวแบบฮิดเดนมาร์คอฟ
- 1.2.3 เพื่อหาวิธีการวิเคราะห์ลักษณะเฉพาะของตัวอักษรธรรมอีสานที่มีประสิทธิภาพ

1.3 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ได้กำหนดขอบเขตของการศึกษาลักษณะของตัวอักษรธรรมอีสานไว้ 91

ตัว จากแบบตัวอักษร 6 แบบ ดังรายละเอียดปรากฏในตารางที่ 1.1

ตารางที่ 1.1 แสดงตัวอักษรธรรมอีสานที่ใช้ในการวิจัย

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
1	ก	๓	๓	๓	๓	๓	๓
2	ข	๔	๔	๔	๔	๔	๔
3	ค	๕	๕	๕	๕	๕	๕

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
4	ฅ	ฅ	ฅ	ฅ	ฅ	ฅ	ฅ
5	ง	ง	ง	ง	ง	ง	ง
6	จ	จ	จ	จ	จ	จ	จ
7	ฉ	ฉ	ฉ	ฉ	ฉ	ฉ	ฉ
8	ช	ช	ช	ช	ช	ช	ช
9	ฌ	ฌ	ฌ	ฌ	ฌ	ฌ	ฌ
10	ญ	ญ	ญ	ญ	ญ	ญ	ญ
11	ฎ	ฎ	ฎ	ฎ	ฎ	ฎ	ฎ
12	ฏ	ฏ	ฏ	ฏ	ฏ	ฏ	ฏ
13	ฑ	ฑ	ฑ	ฑ	ฑ	ฑ	ฑ
14	ฒ	ฒ	ฒ	ฒ	ฒ	ฒ	ฒ
15	ณ	ณ	ณ	ณ	ณ	ณ	ณ
16	ด	ด	ด	ด	ด	ด	ด
17	ต	ต	ต	ต	ต	ต	ต
18	ถ	ถ	ถ	ถ	ถ	ถ	ถ
19	ท	ท	ท	ท	ท	ท	ท
20	ธ	ธ	ธ	ธ	ธ	ธ	ธ
21	น	น	น	น	น	น	น
22	บ	บ	บ	บ	บ	บ	บ
23	ป	ป	ป	ป	ป	ป	ป
24	ผ	ผ	ผ	ผ	ผ	ผ	ผ
25	ฝ	ฝ	ฝ	ฝ	ฝ	ฝ	ฝ

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
26	พ	๓	๓	๓	๓	๓	๓
27	ฟ	๓	๓	๓	๓	๓	๓
28	ภ	๓	๓	๓	๓	๓	๓
29	ม	๓	๓	๓	๓	๓	๓
30	ย	๓	๓	๓	๓	๓	๓
31	ร	๓	๓	๓	๓	๓	๓
32	ล	๓	๓	๓	๓	๓	๓
33	ว	๓	๓	๓	๓	๓	๓
34	ศ	๓	๓	๓	๓	๓	๓
35	ห	๓	๓	๓	๓	๓	๓
36	ฬ	๓	๓	๓	๓	๓	๓
37	อ	๓	๓	๓	๓	๓	๓
38	ฮ	๓	๓	๓	๓	๓	๓
39	ฤา	๓	๓	๓	๓	๓	๓
40	๑	๑	๑	๑	๑	๑	๑
41	๒	๑	๑	๑	๑	๑	๑
42	๓	๑	๑	๑	๑	๑	๑
43	๔	๑	๑	๑	๑	๑	๑
44	๕	๑	๑	๑	๑	๑	๑
45	๖	๑	๑	๑	๑	๑	๑
46	๗	๑	๑	๑	๑	๑	๑
47	๘	๑	๑	๑	๑	๑	๑

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
48	ฮ	๓	๓	๓	๓	๓	๓
49	อ	๐	๐	๐	๐	๐	๐
50	อิ	๒	๒	๒	๒	๒	๒
51	อี	๒	๒	๒	๒	๒	๒
52	อุ	๒	๒	๒	๒	๒	๒
53	อู	๒	๒	๒	๒	๒	๒
54	เอ	๒	๒	๒	๒	๒	๒
55	สส	๓	๓	๓	๓	๓	๓
56	สระ อะ	๓	๓	๓	๓	๓	๓
57	สระ อา	๒	๒	๒	๒	๒	๒
58	สระ อิ	๐	๐	๐	๐	๐	๐
59	สระ อี	๒	๒	๒	๒	๒	๒
60	สระ อื	๑	๑	๑	๑	๑	๑
61	สระ อือ	๑	๑	๑	๑	๑	๑
62	สระ อุ	๒	๒	๒	๒	๒	๒
63	สระ อู	๒	๒	๒	๒	๒	๒
64	เ	๒	๒	๒	๒	๒	๒
65	แ	๒	๒	๒	๒	๒	๒
66	โ	๒	๒	๒	๒	๒	๒
67	สระ อำ	๒	๒	๒	๒	๒	๒
68	ไม้ก	๒	๒	๒	๒	๒	๒
69	นา	๒	๒	๒	๒	๒	๒

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
70	หยอ	ᨠ	ᨠ	ᨠ	ᨠ	ᨠ	ᨠ
71	ไ	ᨡ	ᨡ	ᨡ	ᨡ	ᨡ	ᨡ
72	ใ	ᨢ	ᨢ	ᨢ	ᨢ	ᨡ	ᨡ
73	ย เพื่อง	ᨣ	ᨣ	ᨣ	ᨣ	ᨣ	ᨣ
74	ย	ᨤ	ᨤ	ᨤ	ᨤ	ᨤ	ᨤ
75	นิคหิต	ᨦ	ᨦ	ᨦ	ᨦ	ᨦ	ᨦ
76	น เพื่อง	ᨧ	ᨧ	ᨧ	ᨧ	ᨧ	ᨧ
77	ม เพื่อง	ᨨ	ᨨ	ᨨ	ᨨ	ᨨ	ᨨ
78	ง เพื่อง	ᨩ	ᨩ	ᨩ	ᨩ	ᨩ	ᨩ
79	ล เพื่อง	ᨪ	ᨪ	ᨪ	ᨪ	ᨪ	ᨪ
80	ไม้อัง แตน	ᨫ	ᨫ	ᨫ	ᨫ	ᨫ	ᨫ
81	ร เพื่อง	ᨬ	ᨬ	ᨬ	ᨬ	ᨬ	ᨬ
82	ห เพื่อง	ᨭ	ᨭ	ᨭ	ᨭ	ᨭ	ᨭ
83	ก เพื่อง	ᨮ	ᨮ	ᨮ	ᨮ	ᨮ	ᨮ
84	ฐ เพื่อง	ᨯ	ᨯ	ᨯ	ᨯ	ᨯ	ᨯ
85	จ เพื่อง	ᨰ	ᨰ	ᨰ	ᨰ	ᨰ	ᨰ
86	อ เพื่อง	ᨱ	ᨱ	ᨱ	ᨱ	ᨱ	ᨱ
87	ธ เพื่อง	ᨲ	ᨲ	ᨲ	ᨲ	ᨲ	ᨲ
88	ว เพื่อง	ᨳ	ᨳ	ᨳ	ᨳ	ᨳ	ᨳ
89	ค เพื่อง	ᨴ	ᨴ	ᨴ	ᨴ	ᨴ	ᨴ
90	ไม้ซัด	ᨵ	ᨵ	ᨵ	ᨵ	ᨵ	ᨵ

ที่	ตัวอักษร	ร.ร.สว่างวีระวงศ์ แบบที่ 1	ร.ร.สว่างวีระวงศ์ แบบที่ 2	ม.ราชภัฏ อุบลราชธานี แบบที่ 1	ม.ราชภัฏ อุบลราชธานี แบบ ที่ 2	ม.มหาสารคาม	มจร. วิทยาเขต ขอนแก่น
91	สระอา ยาว	า	า	ว	า	ว	า

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 มีรูปแบบการพิมพ์ตัวอักษรธรรมอีสานบนคอมพิวเตอร์ที่ได้มาตรฐาน
- 1.4.2 ได้วิธีการในการวิเคราะห์ลักษณะเฉพาะของตัวอักษรที่มีประสิทธิภาพ
- 1.4.3 มีระบบและกระบวนการวิธีในการรู้จำตัวอักษรธรรมอีสานที่มีประสิทธิภาพ
- 1.4.4 เพื่อเป็นแนวทางสู่กระบวนการพัฒนาโปรแกรมอ่านภาพอักษรธรรมอีสานที่ปรากฏบนใบลาน หรือสื่ออื่นๆ ต่อไป
- 1.4.5 เป็นหนทางสู่การอนุรักษ์มรดกของชาวอีสาน

1.5 นิยามศัพท์ที่ใช้ในการวิจัย

การวิจัยครั้งนี้ มีคำนิยามศัพท์หรือคำจำกัดความที่กำหนดไว้เป็นการเฉพาะประกอบด้วย
อักษรธรรมอีสาน หมายถึง ตัวอักษรใดๆ ที่เป็นอักษรโบราณซึ่งมักจารึกบนใบลาน เรียกว่า ตัวธรรม หรือ อักษรธรรมอีสาน

รูปแบบตัวอักษรธรรมอีสาน หมายถึง รูปแบบตัวอักษร หรือ Font ของตัวอักษรธรรมอีสานที่มีผู้พัฒนาขึ้นมาเพื่อใช้ในการพิมพ์บนเครื่องคอมพิวเตอร์

รูปภาพอักษร หมายถึง รูปภาพของตัวอักษรของอักษรใดๆ ที่พิมพ์ขึ้นมาแล้วจัดเก็บเป็นไฟล์รูปภาพ

การรู้จำตัวอักษร หมายถึง กระบวนการที่ทำให้เครื่องคอมพิวเตอร์มีความสามารถในการทำงานใกล้เคียงกับมนุษย์ในด้านการรับรู้ภาพอักษรที่ป้อนเข้าไบนั้น เป็นอักษระตัวใด

การลดความหนาอักษร หมายถึง กระบวนการหนึ่งในขั้นตอนของการเตรียมข้อมูลก่อนที่จะนำไปใช้ในระบบการรู้จำตัวอักษร โดยจะทำให้ภาพอักษรมีความหนาลดลงเหลือเพียง 1 จุด (pixel) เท่านั้น แต่ยังคงสัดส่วนและโครงสร้างเดิมไว้

การกระจายความหนาแน่นของจุด หมายถึง การสกัดเอาลักษณะเด่นๆ ของภาพที่ศึกษา โดยกำหนดพื้นที่เป็นสัดส่วนต่างๆ กัน แล้วดูความหนาแน่นของจุดภาพที่กระจายในพื้นที่กำหนด

การกระจายพิกเซลของตัวอักษรแบบไม่เชิงเส้น หมายถึง การปรับความหนาแน่นจุดภาพของตัวอักษรในกรอบพื้นที่ที่กำหนด โดยมีความสัมพันธ์เชิงสมการแบบไม่เชิงเส้น

ตัวแบบฮิดเดนมาร์คอฟ หมายถึง เทคนิคการเรียนรู้ทางสถิติ (Statistical machine learning) ที่มีประสิทธิภาพสูง คิดค้นขึ้นโดยนักคณิตศาสตร์ชาวรัสเซียชื่อ Andrei Andreevich Markov โดยใช้จำแนกรูปแบบของลำดับเหตุการณ์ต่างๆ ซึ่งลำดับของเหตุการณ์จะถูกจำลองให้อยู่ในรูปของลำดับการเปลี่ยนแปลงของสถานะ

