

ชื่อรายงานการวิจัย	การรู้จำอักขระอักษรธรรมอีสานโดยใช้ตัวแบบฮิดเดนมาร์คอฟ
ผู้วิจัย	นายนิรันดร์ เลิศวีรพล, นายสุรศักดิ์ ตั้งสกุล
หน่วยงาน	มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย วิทยาเขตขอนแก่น
ปีงบประมาณ	2549
ทุนอุดหนุนการวิจัย	มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย

บทคัดย่อ

อักขระอักษรธรรมอีสาน เป็นอักขระที่ไม่ใช่มาตรฐานการพิมพ์ทั่วไปบนเครื่องคอมพิวเตอร์ ดังนั้นจึงยากที่จะนำมาเข้าสู่กระบวนการรู้จำ เนื่องจากรูปแบบการเขียนไม่แน่นอน ซับซ้อน และยืดหยุ่นตามลักษณะของภูมิภาคที่ผู้เขียนนั้นอยู่ ในการรู้จำจึงต้องอาศัยการสกัดลักษณะเฉพาะจากลักษณะเด่นหลาย ๆ อย่าง ทั้งนี้เพื่อให้สามารถอธิบายเอกลักษณ์ที่สำคัญของอักขระแต่ละตัวได้อย่างครบถ้วนมากที่สุด งานวิจัยนี้มุ่งเน้นศึกษาการรู้จำอักขระตัวอักษรธรรมอีสานจำนวน 91 ตัวๆ ละ 6 แบบ โดยใช้วิธีการปรับการกระจายของพิกเซลของตัวอักขระแบบไม่เชิงเส้นและคุณลักษณะเด่นของอักขระหลายวิธีร่วมกันโดยใช้ตัวแบบฮิดเดนมาร์คอฟชนิดต่าง ๆ เพื่อศึกษาเปรียบเทียบประสิทธิภาพ โดยมีการกำหนดจำนวนสถานะของตัวแบบในลักษณะที่แน่นอนและไม่แน่นอน ผลที่ได้จากการวิจัยพบว่าจำนวนสถานะที่เหมาะสมและน้อยที่สุดที่ให้ผลอัตราการรู้จำเฉลี่ยสูงสุดสำหรับตัวแบบชนิด Left-Right มีจำนวนสถานะอยู่ระหว่าง 20 - 40 สถานะ โดยมีอัตราการรู้จำเท่ากับ 98.9%

Research Title: The E-Sarn Dharma Character Recognition by Using
Hidden Markov Models

Researcher: Nirandorn Lerdwiraphol and Surasak Tangsakul

Department: Buddhist Research Institute,
Mahachulalongkornrajavidyalaya University

Fiscal Year: 2549 / 2006

Research Scholarship Sponsor: Mahachulalongkornrajavidyalaya University

ABSTRACT

The E-Sarn Dharma Characters are not the standards printing of computer. So, it is very difficult to build up an acceptable algorithm for recognizing them. Because of this, it is of several formalities and flexibilities for writing in other regions. And in the region in which the writer is living. So, in recognition and rememorate it must depend on the prevention of special characteristic from the distinguished characteristics. Thus, in order to be able to explain the importantly special characteristics of each alphabet mostly completely. In this research, we are specific to study and remember the 91 Dhamma Alphabets in each of which is of six forms E-Sarn (Northeastern Provinces).

This study is of the usage of the method to adjust the size of the pixel of the alphabets in the form of non-linear in corporation with a number of the distinctive pattern of E-Sarn Damma Characters to determine the features for the recognition of algorism which utilizes the Hidden Markove Models (HMM), From the comparative study of the effectiveness of the E-Sarn Dhamma Alphabets with the numbers of the statistics of the patterns of the E-Sarn Dhamma Alphabets in the a definite and non-definite characteristics, it is found that the statistical numbers of E-Sarn Dhamma Alphabets are divided into two parts; the proper part and the least one which give the result at the rate of knowing and recognizing at the overate point of 98.9 percent which can be divided at the highest level for the patterns of E-Sarn Dhamma Alphabets of the left-right types, and of the numbers of the E-Sarn Dhamma Alphabets of between 20-40.

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักงานคณะกรรมการวิจัยแห่งชาติ (สวช.) ประจำปีงบประมาณ 2549 ภายใต้แผนงานวิจัยของมหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย ซึ่งได้ดำเนินการวิจัยโดยหัวหน้าฝ่ายห้องสมุดและเทคโนโลยีสารสนเทศ สำนักวิชาการวิทยาเขตขอนแก่น

คณะผู้วิจัยขอขอบคุณคณะผู้บริหารและเจ้าหน้าที่สถาบันวิจัยพุทธศาสตร์ มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย ที่ได้ให้โอกาสในการศึกษาวิจัยครั้งนี้ ซึ่งคณะผู้วิจัยเองก็ได้ทุ่มเททั้งกำลังกายและกำลังใจเพื่อให้ผลงานวิจัยนี้ออกมาดีและมีคุณภาพ สมดังปณิธานของมหาวิทยาลัย

คณะผู้วิจัยขอน้อมกราบพระรัตนตรัยด้วยจิตศรัทธาถึงต่อพระพุทธศาสนา ที่ทำให้งานวิจัยครั้งนี้สำเร็จลุล่วงลงด้วยดี โดยหวังเป็นอย่างยิ่งว่างานวิจัยชิ้นนี้จะเป็นประโยชน์ต่อประเทศชาติและศาสนาในการทำนุบำรุงศิลปวัฒนธรรมในลำดับสืบไป

นายนิรันดร์ เลิศวีรพล และคณะ

20 มีนาคม 2552

สารบัญ

บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญภาพ.....	ช
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	4
1.3 ขอบเขตของการวิจัย.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	9
1.5 นิยามศัพท์ที่ใช้ในการวิจัย.....	9
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	10
2.1 ความเป็นมาของอักษรธรรมอีสาน.....	10
2.2 รูปแบบของตัวอักษรธรรมอีสาน.....	11
2.3 การรู้จำรูปแบบ.....	16
2.4 ชนิดและคุณสมบัติของการรู้จำ.....	16
2.5 กระบวนการรู้จำทางคอมพิวเตอร์.....	18
2.6 ขั้นตอนวิธีลดความหนาของตัวอักษร.....	19
2.7 ตัวแบบฮิดเดนมาร์คอฟ.....	20
2.8 งานวิจัยที่เกี่ยวข้อง.....	24
บทที่ 3 วิธีการดำเนินการวิจัย.....	27
3.1 ข้อมูลแบบตัวอักษรธรรมอีสานที่ใช้ในการวิจัย.....	27
3.2 การแปลงข้อมูล.....	37
3.3 การประมวลผลเบื้องต้น.....	39
3.4 การสกัดลักษณะเฉพาะของตัวอักษรธรรมอีสาน.....	49
3.5 การสร้างตัวแบบ.....	56
3.6 การทดลองและผลการทดลอง.....	61

บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	61
4.1 ผลการวิจัยและอภิปรายผล.....	61
4.2 ผลการทดลอง.....	70
4.3 ปัญหาต่างๆ ที่ส่งผลกระทบต่อประสิทธิภาพการรู้จำ.....	105
บทที่ 5 สรุป อภิปรายผลและข้อเสนอแนะ.....	108
5.1 สรุปผลการวิจัย.....	108
5.2 อภิปรายผลการวิจัย.....	112
5.3 ข้อเสนอแนะ.....	113
บรรณานุกรม.....	115
ภาคผนวก.....	117
ผนวก ก การเตรียมข้อมูลภาพอักษรธรรมอีสาน.....	118
ประวัติผู้วิจัยและคณะ.....	128

สารบัญตาราง

ตารางที่

1.1 แสดงตัวอักษรและอักขรธรรมอีสานที่ใช้ในการวิจัย.....	4
3.1 ค่าความน่าจะเป็นของการเปลี่ยนสถานะของตัวแบบชนิด Left-Right.....	57
4.1 ผลการรู้จำของตัวอักษรธรรมอีสาน.....	65
5.1 แสดงตัวอย่างค่าความน่าจะเป็นของลำดับการสังเกตของตัวอักษร สระอำ.....	74
5.2 แสดงตัวอย่างค่าความน่าจะเป็นของลำดับการสังเกตของตัวอักษร ก.....	74

สารบัญภาพ

ภาพที่

1.1 แสดงอักขระอักษรธรรมอีสานตัวอักษร ช.....	2
1.2 ตัวอย่างลายมือเขียนอักษรธรรมอีสานของพระพุทฺธิสารมุนี.....	3
1.3 ตัวอย่างลายมือเขียนอักษรธรรมอีสานของพระพุทฺธิสารมุนี.....	3
2.1 แสดงการกำเนิดตัวอักษรตามแนวของ ก่องแก้ว วีระประจักษ์.....	12
2.2 แสดงระดับของอักษรธรรมอีสาน.....	17
2.3 กระบวนการรู้จำ.....	20
2.4 Neighborhood ของ N(p).....	22
2.5 แสดงขั้นตอนการสอนตัวแบบ.....	24
2.6 Block Diagram การรู้จำอักขระโคใช้ตัวแบบฮิดเดนมาร์คอฟ.....	26
2.7 ก จำนวนสถานะที่เหมาะสมของตัวแบบชนิด Left-Right ของอักขระกลุ่มที่ 1.....	27
2.7 ข จำนวนสถานะที่เหมาะสมของตัวแบบชนิด Left-Right ของอักขระกลุ่มที่ 2.....	28
3.1 แสดงตัวอย่างเว็บไซต์โรงเรียนสว่างวีรวงศ์ในส่วนของการเรียนอักษรธรรมอีสาน.....	29
3.2 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ Tham2004.....	30
3.3 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ Thamawatra.....	31
3.4 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ UbWManut.....	32
3.5 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ UbWThawat.....	33
3.6 แสดงตัวอย่างเว็บไซต์เรียนอักษรโบราณอีสานของมหาวิทยาลัยราชภัฏอุบลราชธานี.....	34
3.7 แสดงตัวอย่างหน้าเว็บโครงการอนุรักษ์โบราณในภาคตะวันออกเฉียงเหนือ.....	35
3.8 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ Agsomtham. Tang.....	36
3.9 แสดงตัวอย่างแบบอักษรธรรมอีสานชื่อ AksornDharm2.....	37
3.10 แสดงขั้นตอนการแปลงข้อมูล.....	38
3.11 แสดงตัวอย่างการพิมพ์อักขระด้วยโปรแกรม Paint.....	38
3.12 แสดงภาพอักขระอักษรธรรมอีสาน (ก) ก่อนแปลง และ (ข) หลังแปลง.....	39
3.13 ขั้นตอนกระบวนการเตรียมข้อมูล.....	40
3.14 แสดงกฎการตรวจหาขอบภาพ.....	41
3.15 แสดงกฎการจัดส่วนเกิน.....	41
3.16 แสดงทิศทางของการพิจารณากำหนดกฎ.....	42
3.17 ผลลัพธ์จากกฎการตรวจหาขอบภาพ.....	43

3.18 แสดงส่วนเกินที่ต้องถูกกำจัด.....	44
3.19 แสดงภาพต้นฉบับ(ก) และผลลัพธ์ที่ได้(ข) เมื่อผ่านกระบวนการลดความหนา.....	44
3.20 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม Tham2004.....	45
3.21 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม Thamawatra.....	45
3.22 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม UbWManut.....	46
3.23 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม UbWThawat.....	46
3.24 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม Agsorntham.Tang.....	47
3.25 แสดงผลการลดความหนาของอักขระแบบอักษรธรรม AksornDharm2.....	48
3.26 ตัวอย่างการปรับการกระจายความหนาแน่นของจุดของอักขระ “ ๒๒ ”.....	49
3.27 การแบ่งโซนของภาพ.....	50
3.28 3×3 neighborhoods.....	53
3.29 ลักษณะการวัดความสูงและความยาวของอักขระ.....	55
3.30 ลักษณะของตัวแบบชนิด Left-Right.....	57
3.31 ลักษณะของสัญลักษณ์ในแต่ละสถานะของตัวแบบ.....	58
3.32 ขั้นตอนการสอนและประมาณค่าตัวแปรใหม่ให้กับตัวแบบ.....	60
3.33 ขั้นตอนการสร้างตัวแบบของการทดลองที่ 2.....	62
4.1 ผลลัพธ์ของการสอนตัวแบบของตัวอักขระ “ ๒๒ ”.....	63
4.2 ค่าความน่าจะเป็นของลำดับการสังเกตต่อตัวแบบของตัวอักขระ “ ๒๒ ”.....	64
4.3 จำนวนสถานะที่เหมาะสมของตัวแบบตัวอักษรธรรมอีสาน.....	65
4.4 แสดงตัวอักขระสระอาที่แตกต่างกัน โดยภาพ (ก) มีหัว และภาพ (ข) ไม่มีหัว.....	71
4.5 แสดงตัวอักขระสระอาที่ทำ Thinning.....	71
4.6 แสดงตัวอักขระสระอาที่ปรับการกระจายของจุด.....	71
4.7 แสดงค่าความน่าจะเป็นของลำดับการสังเกตตัวอักขระ “ ๒ ”.....	72
5.1 แสดงขั้นตอนการสกัดลักษณะเฉพาะของอักษรธรรมอีสาน.....	73