

ระบบออนไลน์สำหรับการรู้จำตัวพิมพ์อักษรไทยและตัวพิมพ์อักษรอังกฤษ



นายนิธิพัฒน์ ชีชวาลพาณิชย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2537

ISBN 974-584-708-9

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I14083035

AN ON-LINE RECOGNITION SYSTEM FOR PRINTED THAI CHARACTERS
AND PRINTED ENGLISH CHARACTERS

MR. NITIPAT CHADCHAVALPANICHAYA

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Computer Engineering

Graduate School
Chulalongkorn University

1994


ISBN 974-584-708-9

หัวข้อวิทยานิพนธ์
โดย
ภาควิชา
อาจารย์ที่ปรึกษา

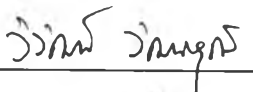
ระบบออนไลน์สำหรับการรู้จำตัวพิมพ์อักษรไทยและตัวพิมพ์อักษรอังกฤษ
นายนิธิพัฒน์ ชัชวาลพาณิชย์
วิศวกรรมคอมพิวเตอร์
อาจารย์ ชัยศิริ ปิณฑิตานนท์
ผู้ช่วยศาสตราจารย์ ดร. วีระ วีวิทัทธ





บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้วิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

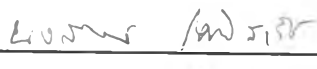

คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ทวาร วัชรราษฎร์)

คณะกรรมการสอบวิทยานิพนธ์


ประธานกรรมการ
(อาจารย์ วีวัฒน์ วัฒนาวุฒิ)


อาจารย์ที่ปรึกษา
(อาจารย์ ชัยศิริ ปิณฑิตานนท์)


อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร. วีระ วีวิทัทธ)


กรรมการ
(อาจารย์ นงลักษณ์ โคควาวิสารักษ์)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

นิพนธ์ ชัชวาลพาณิชย์ : ระบบออนไลน์สำหรับการรู้จำตัวพิมพ์อักษรไทยและตัวพิมพ์
อักษรอังกฤษ (AN ON-LINE RECOGNITION SYSTEM FOR PRINTED THAI
CHARACTERS AND PRINTED ENGLISH CHARACTERS)

อาจารย์ที่ปรึกษา : อ.ชัยศิริ ปัดพิตานนท์, ผศ.ดร.วีระ ธีรพิทักษ์, 117 หน้า.

ISBN 974-584-708-9



วิทยานิพนธ์นี้มีจุดประสงค์ เพื่อให้เครื่องคอมพิวเตอร์สามารถรู้จำตัวพิมพ์อักษรไทยได้
3 ระดับ และตัวพิมพ์อักษรอังกฤษ โดยจะนำผลที่ได้ (รหัสตัวอักษร) ไปเก็บไว้ในแฟ้มข้อมูลตัวอักษรที่
สามารถนำมาใช้ประโยชน์ต่อไป

วิทยานิพนธ์นี้แบ่งเป็น 4 ขั้นตอนใหญ่ ๆ ดังนี้

ขั้นตอนที่ 1 การวิเคราะห์เส้นแสดงขอบของตัวอักษร

ขั้นตอนที่ 2 การหาลักษณะสำคัญของตัวอักษร

ขั้นตอนที่ 3 เปรียบเทียบตัวอักษรที่ทำการรู้จำกับตัวอักษรต้นแบบ

ขั้นตอนที่ 4 นำรหัสของตัวอักษรต้นแบบที่ได้มาเก็บไว้ในแฟ้มข้อมูลตัวอักษร

ผลการวิจัยปรากฏว่า ความถูกต้องในการรู้จำประมาณ 90.09 เปอร์เซ็นต์ ใช้เวลาใน
การรู้จำประมาณ 10.26 วินาทีต่อ 1 ตัวอักษร

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2536

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



C317175 : MAJOR COMPUTER SCIENCE

KEY WORD: CHARACTER RECOGNITION/THAI CHARACTER/ENGLISH CHARACTER

NITIPAT CHADCHAVALPANICHAYA : AN ON-LINE RECOGNITION SYSTEM FOR PRINTED THAI CHARACTERS AND PRINTED ENGLISH CHARACTERS.

THESIS ADVISOR : MR. CHAISIRI PANDHITANONT,

ASST. PROF. WEERA RIEWPITUK, D.Ing. 117 pp. ISBN 974-584-708-9

The purpose of this thesis is for computer can recognize Thai character for 3 levels and also English character. By taking the result (ASCII code) to save in the text file which can be used in any application.

This thesis provides 4 steps :

1. Contour analysis
2. Finding the important character characteristic
3. Comparison the character which recognize with original character
4. Taking the original ASCII code to save in text file.

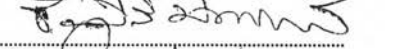
The result of research appears that the accuracy for recognition about 90.09 % and spend the time for recognized about 10.26 seconds per 1 character.

ภาควิชา.....วิศวกรรมคอมพิวเตอร์

สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา.....2536

ลายมือชื่อนิสิต.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ ด้วยความช่วยเหลืออย่างดียิ่งของ อาจารย์ ชัยศิริ ปันพิदानนท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ และผู้ช่วยศาสตราจารย์ ดร. วีระ รวีพิทักษ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งท่านได้กรุณาสละเวลาให้คำแนะนำ และ ช่างคิดเห็นต่าง ๆ ในการวิจัยมาด้วยดีตลอด ผู้วิจัยจึงขอกราบขอบพระคุณในความกรุณาเป็นอย่างสูง

ขอขอบพระคุณ คณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ที่ได้กรุณาให้คำแนะนำ และช่วยเหลือในด้านต่าง ๆ

ขอขอบพระคุณ ท่านคณะกรรมการสอบวิทยานิพนธ์เป็นอย่างสูง ที่ได้ช่วยพิจารณาให้คำแนะนำ ตรวจสอบ แก้ไข และอนุมัติวิทยานิพนธ์ฉบับนี้

รวมทั้งขอขอบพระคุณเพื่อน ๆ และทุกท่านที่ได้ให้ความช่วยเหลือ และกำลังใจแก่ผู้วิจัยตลอดมา

ทำนนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ที่ได้สนับสนุนในด้านการเงิน และน้อม ๆ ที่คอยให้กำลังใจและช่วยเหลือ แก่ผู้วิจัยเสมอมาจนสำเร็จจนสำเร็จการศึกษา



สารบัญ

หน้า

| | |
|--|---|
| บทคัดย่อภาษาไทย | ง |
| บทคัดย่อภาษาอังกฤษ | จ |
| กิตติกรรมประกาศ | ฉ |
| สารบัญ | ช |
| สารบัญรูปภาพ | ญ |
| สารบัญรูปตาราง | ฎ |
| บทที่ | |
| 1. บทนำ | 1 |
| 1.1 ความเป็นมาของปัญหา | 1 |
| 1.2 วัตถุประสงค์ | 2 |
| 1.3 ขอบเขตของการวิจัย | 2 |
| 1.4 ขั้นตอนการวิจัย | 3 |
| 1.5 ประโยชน์ที่คาดว่าจะได้รับ | 3 |
| 2. ทฤษฎีแนวคิดที่สำคัญ | 4 |
| 2.1 การรู้จำตัวอักษร | 4 |
| 2.2 การรู้จำด้วยระบบออนไลน์ | 4 |
| 2.3 การรู้จำด้วยระบบออฟไลน์ | 4 |
| 2.4 ตัวอักษรที่นำมาประยุกต์ใช้ในการรู้จำ | 5 |
| 2.5 เทคนิคที่ใช้ในการรู้จำตัวอักษร | 5 |
| 2.5.1 ลักษณะทางโกลบอล | 5 |
| 2.5.2 ลักษณะทางโลคอล | 6 |
| 2.6 ผลของการวิจัยการรู้จำตัวอักษรที่ผ่านมา | 8 |

| | | |
|-------|--|----|
| 3. | การออกแบบ และ พัฒนาโปรแกรม | 10 |
| 3.1 | ขั้นตอนการเขียนรู้ | 10 |
| 3.1.1 | รับข้อมูลตัวพิมพ์อักษรไทยและตัวพิมพ์อักษรอังกฤษ .. | 11 |
| 3.1.2 | กำจัดสัญญาณรบกวนและแก้ไขเพิ่มเติมเส้นตรง | 11 |
| 3.1.3 | หาจุดเริ่มต้นของตัวอักษร | 40 |
| 3.1.4 | วิเคราะห์เส้นแสดงขอบของตัวอักษร | 40 |
| 3.1.5 | กำหนดค่าลักษณะสำคัญของตัวอักษร | 44 |
| 3.1.6 | กำหนดกลุ่มของตัวอักษรต้นแบบ | 51 |
| 3.1.7 | เก็บข้อมูลไว้ในแฟ้มข้อมูลตัวอักษรต้นแบบ | 55 |
| 3.2 | ขั้นตอนการรู้จำ | 60 |
| 3.2.1 | รับข้อมูลตัวพิมพ์อักษรไทยและตัวพิมพ์อักษรอังกฤษ .. | 60 |
| 3.2.2 | กำจัดสัญญาณรบกวนและแก้ไขเพิ่มเติมเส้นตรง | 60 |
| 3.2.3 | หาจุดเริ่มต้นของตัวอักษร | 60 |
| 3.2.4 | วิเคราะห์เส้นแสดงขอบของตัวอักษร | 61 |
| 3.2.5 | กำหนดค่าลักษณะสำคัญของตัวอักษร | 61 |
| 3.2.6 | กำหนดกลุ่มตัวอักษรต้นแบบ | 61 |
| 3.2.7 | เปรียบเทียบกับตัวอักษรต้นแบบในกลุ่มตัวอักษร | 61 |
| 3.2.8 | จัดเก็บข้อมูลที่ได้ | 69 |
| 3.3 | การหาส่วนโค้งและรูปแบบของส่วนโค้ง | 71 |
| 3.3.1 | การหาส่วนโค้ง | 71 |
| 3.3.2 | การหารูปแบบของส่วนโค้ง | 74 |
| 3.4 | การหาเส้นตรง | 80 |
| 3.5 | การจัดระดับตัวอักษร | 80 |
| 4. | ผลการวิจัย | 89 |
| 4.1 | อุปกรณ์ที่ใช้ในการวิจัย | 89 |
| 4.2 | จำนวนตัวอักษรที่ใช้ในการวิจัย | 90 |
| 4.3 | ตัวอักษรที่ใช้ในการวิจัย | 90 |
| 4.4 | ผลการรู้จำตัวอักษร | 91 |

| | |
|----------------------------------|-----|
| 5. บทสรุปและข้อเสนอแนะ | 100 |
| 5.1 สรุปผลการวิจัย | 100 |
| 5.2 ข้อเสนอแนะ | 107 |
| เอกสารอ้างอิง | 111 |
| ภาคผนวก | 113 |
| ภาคผนวก ก | 114 |
| ภาคผนวก ข | 115 |
| ประวัติผู้เขียนวิทยานิพนธ์ | 117 |

สารบัญรูปภาพ

| รูปที่ | หน้า |
|--|------|
| 3.1 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 1 | 12 |
| 3.2 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 2 | 13 |
| 3.3 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 3 | 14 |
| 3.4 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 4 | 15 |
| 3.5 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 5 | 16 |
| 3.6 แสดงการกำจัดสัญญาณรบกวนรูปแบบที่ 6 | 17 |
| 3.7 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 1 | 19 |
| 3.8 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 2 | 20 |
| 3.9 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 3 | 21 |
| 3.10 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 4 | 22 |
| 3.11 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 5 | 23 |
| 3.12 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 6 | 24 |
| 3.13 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 7 | 25 |
| 3.14 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 8 | 26 |
| 3.15 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 9 | 27 |
| 3.16 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 10 | 28 |
| 3.17 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 11 | 29 |
| 3.18 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 12 | 30 |
| 3.19 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 13 | 31 |
| 3.20 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 14 | 32 |
| 3.21 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 15 | 33 |
| 3.22 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 16 | 34 |
| 3.23 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 17 | 35 |
| 3.24 แสดงเมตริกซ์ของการแก้ไขเส้นตรงรูปแบบที่ 18 | 36 |
| 3.25 แสดงตัวอย่างการกำจัดสัญญาณรบกวนและแก้ไขเพิ่มเติมเส้นตรง และ เส้นแสดงขอบของตัวอักษรที่หาได้ | 37 |

| | | |
|------|---|-----|
| 3.26 | ผังงานแสดงการกำจัดสัญญาณรบกวนและแก้ไขเส้นตรง | 38 |
| 3.27 | แสดงเมตริกซ์การตรวจสอบจุดต่อไปของเส้นแสดงขอบตัวอักษร | 41 |
| 3.28 | แสดงการวิเคราะห์เส้นแสดงขอบของตัวอักษร | 42 |
| 3.29 | ผังงานแสดงการวิเคราะห์เส้นแสดงขอบของตัวอักษร | 43 |
| 3.30 | แสดงการหาตำแหน่งของหัวของตัวอักษร | 46 |
| 3.31 | แสดงการหาระยะห่างระหว่างหัวของตัวอักษร | 47 |
| 3.32 | ผังงานแสดงการกำหนดค่าลักษณะสำคัญของตัวอักษร | 50 |
| 3.33 | แสดงการเก็บเพิ่มข้อมูลทั้ง 5 รูปแบบ | 57 |
| 3.34 | ผังงานแสดงการเก็บข้อมูลไว้ในแฟ้มข้อมูลตัวอักษรต้นแบบ | 58 |
| 3.35 | แสดงขั้นตอนการทำงานการเรียนรู้ตัวอักษร | 59 |
| 3.36 | ผังงานแสดงการเปรียบเทียบกับตัวอักษรต้นแบบ | 68 |
| 3.37 | แสดงขั้นตอนการทำงานการรู้จำตัวอักษร | 70 |
| 3.38 | แสดงตัวอย่างการหามุมของส่วนโค้ง | 72 |
| 3.39 | แสดงตัวอย่างการหาส่วนโค้ง | 73 |
| 3.40 | แสดงการหารูปแบบของส่วนโค้ง | 74 |
| 3.41 | แสดงส่วนโค้งรูปแบบที่ 1 | 75 |
| 3.42 | แสดงส่วนโค้งรูปแบบที่ 2 | 76 |
| 3.43 | แสดงส่วนโค้งรูปแบบที่ 3 | 77 |
| 3.44 | แสดงส่วนโค้งรูปแบบที่ 4 | 78 |
| 3.45 | ผังงานแสดงการหาส่วนโค้ง และรูปแบบของส่วนโค้ง | 79 |
| 3.46 | แสดงตัวอักษรที่จะกำหนดแถวและสดมภ์ | 82 |
| 3.47 | แสดงการกำหนดแถวและสดมภ์ให้ตัวอักษร | 83 |
| 3.48 | แสดงตัวอย่างของการตรวจสอบว่าตัวอักษรอยู่บรรทัดเดียวกัน หรือไม่ | 85 |
| 3.49 | ผังงานแสดงการจัดระดับตัวอักษร | 87 |
| 4.1 | แสดงตัวอักษรที่จะทำการรู้จำก่อนทำการกำจัดสัญญาณรบกวน | 92 |
| 4.2 | แสดงตัวอักษรที่จะทำการรู้จำหลังทำการกำจัดสัญญาณรบกวน | 93 |
| 5.1 | แสดงการกำจัดสัญญาณรบกวนที่คิค้นใหม่รูปแบบที่ 1 | 109 |
| 5.2 | แสดงการกำจัดสัญญาณรบกวนที่คิค้นใหม่รูปแบบที่ 2 | 110 |

สารบัญตาราง

| ตารางที่ | หน้า |
|----------|--|
| 4.1 | แสดงจำนวนตัวอักษรที่ใช้ในการรู้จำตัวอักษร 97 |
| 5.1 | แสดงผลสรุปการวิจัยการรู้จำตัวพิมพ์อักษร 102 |
| 5.2 | แสดงการรู้จำตัวอักษรรูปแบบเดียวในระบบออฟไลน์ 104 |
| 5.3 | แสดงการรู้จำตัวอักษรหลายรูปแบบในระบบออฟไลน์ 105 |
| 5.4 | แสดงการรู้จำตัวอักษรในระบบออนไลน์ 106 |