การรู้จำสายอักขระไทยตัวพิมพ์โดยวิธีซินแทกติก

นาย สมศักดิ์ คงถาวรวัฒนา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2539

ISBN 974-634-368-8

RECOGNITION OF THAI PRINTED CHARACTER STRING BY THE SYNTACTIC METHOD

Mr Somsak Kongthawornwattana

A Thesis Submitted in Partial Fulfillment of Requirements

for the Degree of Master of Engineering

Department of Electrical

Graduate School

Chulalongkorn University

1996

ISBN 974-634-368-8

หัวข้อวิทยานิพนธ์	การรู้จำสายอักขระไทยตัวพิมพ์โยวิธีชินแทกติก
โดย	นาย สมศักดิ์ คงวรวัฒนา
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร. สันติ ถุงสุวรรณ)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณียนาวิน)

.....อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล)

.....กรรมการ
(ดร. วิวัฒน์ วงศ์วราวิภาร์)

.....กรรมการ
(ดร. บวรกุล จิตต์ประเสริฐ)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

สมศักดิ์ คงถาวรวัฒนา : การรู้จำสายอักขระไทยกัวพิมพ์โดยวิธีซินแทกติก(RECOGNITION of Thai-Printed character string by the Syntactic Method) อ. ที่ปรึกษา : รศ. ดร. สมชาย จิตะพันธ์กุล, 69 หน้า. ISBN 974-634-368-8

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อหาอัลกอริทึมต้นแบบที่สามารถนำไปใช้การรู้จำสายอักษรตัวพิมพ์ไทยเพื่อให้ผลลัพธ์ที่ได้ออกมาเป็นลักษณะของแฟ้มข้อมูลของตัวอักษร ตามมาตรฐานภาษาไทยของ สมอ. โดยตัวอักษรที่จะนำมาทดสอบจะเป็นตัวอักษรแบบ EucrosiaUPC ขนาด 18 points

สำหรับระบบการรู้จำสายอักษรตัวพิมพ์ไทยประกอบด้วยขั้นตอนต่าง ๆ ที่สำคัญคือ ขั้นตอนการแยกกลุ่มของข้อมูลภาพโดยใช้เทคนิคการหาขอบภาพ การจัดเรียงตัวอักษรโดยใช้วิธีการพิจารณาเส้นฐานและขนาดตัวอักษร ขั้นตอนการรู้จำตัวอักษรโดยใช้วิธีซินแทกติก (สนธยา, 2537)

ผลการทดลองซึ่งใช้สายตัวอักษร 150 สายรวม 1,974 ตัวอักษร พบว่าเมื่อนำภาพตัวอักษรที่ต้องการรู้จำมาทดสอบกับระบบการรู้จำที่พัฒนาขึ้นนั้น สามารถที่จะรู้จำสายอักษรเหล่านั้นได้ โดยผลของการรู้จำสำหรับการวิจัยนี้มีอัตราการรู้จำ 92.70% ไม่สามารถรู้จำได้ 2.90% และรู้จำผิด 4.40%

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต [illegible]

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา [illegible]

ปีการศึกษา 2538 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

C515692 : MAJOR ELECTRICAL ENGINEERING
KEY WORD: RECOGNITION/SYSTACTIC/THAI-PRINTED CHARACTER

SOMSAK KONGTHAWORNWATTANA : RECOGNITION OF THAI-PRINTED CHARACTER STRING BY THE SYNTACTIC METHOD. THESIS ADVISOR : ASSO. PROF. SOMCHAI JITAPUNKUL. Ph.D. 69pp. ISBN 974-634-368-8

This thesis proposed to use the syntactic method to recognize Thai-printed character string. The process composed of 3 steps, image segmentation based on edge detection technique, character sorting obtained from the determination of base line and the size of characters and lastly the character recognition process using the syntactic method (Sonthaya 1994). The input file was limited to EucrosiaUPC font of 18 points only.

From the experiment using 150 character strings with 1,974 characters in total, resulted in 92.70%, 4.4%, and 2.9% of correct, wrong, and undecided recognition rates respectively.

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2538 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ของการวิจัยมาด้วยดีตลอด และ คุณชนิษฐา พรายงาม ที่มีส่วนช่วยเหลือในการให้คำปรึกษาการเขียนโปรแกรมภาษา C และในการจัดพิมพ์วิทยานิพนธ์นี้จนสำเร็จลุล่วงไปด้วยดี

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งสนับสนุนในด้านการเงินและให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า