ระบบรู้จำอักษรภาษาไทย
โดยใช้ลักษณะบ่งความต่างของตัวอักษรไทย

นาย วิชา พานิช

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2539

ISBN 974-635-567-8

A THAI CHARACTER RECOGNITION SYSTEM BASED ON DISTINCTIVE FEATURES OF THAI CHARACTERS

Mr. Wicha Panich

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering

Department of Electrical Engineering

Graduate School

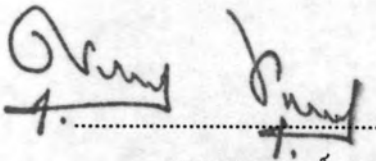
Chulalongkorn University

Academic Year 1996

ISBN 974-635-567-8

หัวข้อวิทยานิพนธ์	ระบบรู้จำอักษรภาษาไทยโดยใช้ลักษณะบ่งความต่างของตัวอักษรไทย
โดย	นาย วิชา พานิช
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล

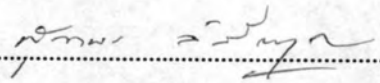
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต



.....รักษาราชการแทนคณบดีบัณฑิตวิทยาลัย

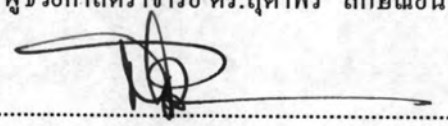
(ศาสตราจารย์ นายแพทย์ ศุภวัฒน์ ชุติวงศ์)

คณะกรรมการสอบวิทยานิพนธ์



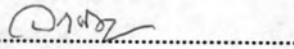
..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.สุดาพร ลักษณียนาวิน)



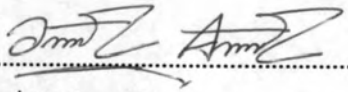
..... อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)



..... กรรมการ

(อาจารย์ ดร.วาทิต เบญจพลกุล)



..... กรรมการ

(ดร. จุฬารัตน์ ตันประเสริฐ)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

วิชา พานิช : ระบบรู้จำอักษรภาษาไทยโดยใช้ลักษณะบ่งความต่างของตัวอักษรไทย (A THAI CHARACTER RECOGNITION SYSTEM BASED ON DISTINCTIVE FEATURES OF THAI CHARACTERS) อ.ที่ปรึกษา : รศ.ดร.สมชาย จิตะพันธ์กุล , 82 หน้า. ISBN 974-635-567-8

วิทยานิพนธ์ฉบับนี้มีจุดมุ่งหมายเพื่อสร้างระบบรู้จำอักษรภาษาไทยโดยใช้ลักษณะบ่งความต่างของอักษรไทย ซึ่งประกอบด้วยงาน 3 ส่วนหลักคือ ส่วนรู้จำอักษรเดี่ยว ส่วนแยกอักษรที่ติดกัน และ ส่วนวิเคราะห์เอกสาร

ในส่วนการรู้จำตัวอักษรภาษาไทยใช้การแบ่งกลุ่มโดยใช้ลักษณะของโครงสร้างหลักร่วมกับระดับของอักษรโดยแบ่งเป็นอักษรระดับบน 1 กลุ่ม ระดับล่าง 1 กลุ่ม และระดับกลางอีก 7 กลุ่ม แล้วจึงแยกแยะในกลุ่มย่อยโดยใช้ลักษณะบ่งความต่างของอักษรไทย ในส่วนการตัดแยกอักษรที่ติดกันนั้นใช้ลักษณะบ่งความต่างของอักษรไทยแบ่งประเภทของการติดกันโดยใช้ระดับของอักษรได้เป็น 10 กลุ่มแล้วใช้วิธีเฉพาะของแต่ละกลุ่มในการตัดแยก ในส่วนการวิเคราะห์เอกสารมีการแก้ความเอียงของเอกสาร การแยกคอลัมน์และแยกบรรทัดตัวอักษร

โดยทำการทดสอบบนเครื่องไมโครคอมพิวเตอร์ CPU 80486DX2-80 กับอักษรกว่า 50,000 ตัวอักษร ได้ผลการรู้จำร้อยละ 97.6 และใช้เวลาเฉลี่ยในการรู้จำ 36.4 อักษรต่อวินาที

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2539 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

C815897 : MAJOR ELECTRICAL ENGINEERING
KEY WORD: CHARACTER RECOGNITION / THAI DISTINCTIVE FEATURES

WICHA PANICH : A THAI CHARACTER RECOGNITION SYSTEM BASED ON DISTINCTIVE FEATURES OF THAI CHARACTERS. THESIS ADVISOR : ASSO. PROF. SOMCHAI JITAPUNKUL, Dr.Ing. 82 pp . ISBN 974-635-567-8

The objective of this thesis is to create a Thai character recognition system based on Thai distinctive features that consists of 3 main parts : a recognition module of single character , a module of segmentation of connected characters image , and a document analysis module.

In the recognition module of single character , the primary structures and the level of the character are used to classify thai characters into 9 groups . They composed of one upper level group , one lower level group and seven middle level groups . Distinctive features of Thai characters are used to classify the member in each groups. In the part of segmentation of connected characters , Thai distinctive features are used to identify the group of connected character . We have 10 groups and 10 methods to segment that connected characters to be single character. The document analysis module provides an algorithm to deskew , detect the columns , detect the lines of the scanned document to create a text file of the character strings in each column and each line.

By using microcomputer of CPU 80486DX2-80 to test the document that contains about 50,000 characters, recognition rate is 97.6% . The average processing time is 36.4 characters per second.

ภาควิชา วิศวกรรมสำรวจ ลายมือชื่อนิสิต วิชา พานิช

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา Somchai Jitapunkul

ปีการศึกษา 2539 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ของการวิจัยมาด้วยดีตลอด ขอขอบคุณทุกท่านที่ไม่ได้เอ่ยนามในที่นี้ ที่ช่วยให้ข้อคำแนะนำและให้กำลังใจเสมอมา

สุดท้ายนี้ ผู้วิจัยใคร่ขอกราบของพระคุณบิดา มารดา ซึ่งได้ให้การสนับสนุนและให้กำลังใจแก่ผู้วิจัยเสมอจนสำเร็จการศึกษา

สารบัญ

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

สารบัญภาพ (ต่อ)

หน้า

สารบัญภาพ (ต่อ)

หน้า