การประยุกต์ใช้การโปรแกรมตรรกะเชิงอุปนัยในการรู้จำตัวพิมพ์อักษรภาษาไทย

นางสาว อภิญญา สุพรรณวรรษา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2540

ISBN 974-637-943-7

AN APPLICATION OF INDUCTIVE LOGIC PROGRAMMING TO THAI PRINTED CHARACTER RECOGNITION

Miss Apinya Supanwassa

A Thesis Submitted in Partial Pulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
Graduate School
Chulalongkorn University
Academic Year 1997

ISBN 974-637-943-7

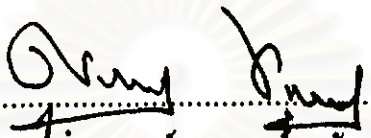
หัวข้อวิทยานิพนธ์ การประยุกต์ใช้การโปรแกรมตรรกะเชิงอุปนัยในการรู้จำตัวพิมพ์อักษรภาษาไทย

โดย นางสาว อภิญญา สุพรรณวรรษา

ภาควิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษา อาจารย์ ดร. บุญเสริม กิจศิริกุล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต



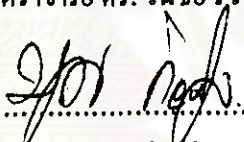
.....คณบดีบัณฑิตวิทยาลัย

(ศาสตราจารย์ นายแพทย์ ศุภวัฒน์ ชุติวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

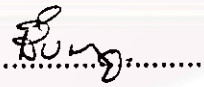
.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. วันชัย ริ้วไพบูลย์)



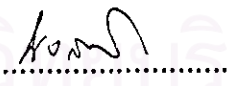
.....อาจารย์ที่ปรึกษา

(อาจารย์ ดร. บุญเสริม กิจศิริกุล)



.....กรรมการ

(อาจารย์ ดร. สืบสกุล พิภพมงคล)



.....กรรมการ

(ผู้ช่วยศาสตราจารย์ นงลักษณ์ โควาวิสารัช)

อภิญญา สุพรรณวรรษา : การประยุกต์ใช้การโปรแกรมตรรกะเชิงอุปนัยในการรู้จำตัวพิมพ์อักษรภาษาไทย (AN APPLICATION OF INDUCTIVE LOGIC PROGRAMMING TO THAI PRINTED CHARACTER RECOGNITION) อ.ที่ปรึกษา : อาจารย์ ดร. บุญเสริม กิจศิริกุล , 78 หน้า. ISBN 974-637-943-7.

งานวิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ เพื่อประยุกต์ใช้การโปรแกรมตรรกะเชิงอุปนัย หรือ ไอแอลพี ในการรู้จำตัวพิมพ์อักษรภาษาไทย ระบบไอแอลพีที่เลือกใช้ คือ ระบบ PROGOL ขั้นตอนการวิจัยเริ่มจากการใช้ PROGOL ในการเรียนรู้ตัวพิมพ์อักษรภาษาไทย ซึ่งข้อมูลที่จำเป็นต่อการเรียนรู้ ได้แก่ ตัวอย่าง และ ความรู้ส่วนหลัง ผลที่ได้จากการเรียนรู้ คือ กลุ่มของกฎ ซึ่งแต่ละกฎจะนิยามลักษณะสำคัญของตัวอักษรภาษาไทยแต่ละตัว ขั้นตอนถัดมา คือ การนำกฎที่ได้จากการเรียนรู้มาใช้ในการรู้จำตัวอักษร โดยเปรียบเทียบกฎที่ได้กับตัวอักษรที่ต้องการรู้จำ และเลือกกฎที่ตรงกับลักษณะของตัวอักษรนั้นๆมากที่สุดให้เป็นผลการรู้จำ การทดสอบวิธีการดังกล่าวแบ่งออกเป็น 2 การทดลอง การทดลองแรกเพื่อทดสอบการรู้จำตัวอักษรในรูปแบบที่ไม่เคยเรียนรู้มาก่อน โดยใช้ตัวอักษรรูปแบบ EUCROSIA ในการเรียนรู้ และ ใช้รูปแบบ CORDIA ในการทดสอบการรู้จำ พบว่าผลการรู้จำมีความถูกต้อง 87.38 % จากจำนวนตัวอักษรที่ทำการทดสอบ 539 ตัวอักษร การทดลองที่สองเพื่อทดสอบการรู้จำตัวอักษรที่มีสัญญาณรบกวน โดยใช้ตัวอักษรรูปแบบ CORDIA และ EUCROSIA ในการเรียนรู้ และ นำตัวอักษรทั้งสองรูปแบบนั้นไปคัดลอกด้วยเครื่องถ่ายเอกสารได้จำนวนตัวอักษร 2156 ตัวอักษร แล้วจึงนำมาทดสอบการรู้จำ พบว่าผลการรู้จำมีความถูกต้อง 87.89 % เวลาในการรู้จำโดยเฉลี่ย 0.13 วินาทีต่อการรู้จำ 1 ตัวอักษร

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต อภิญญา สุพรรณวรรษา
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์	ลายมือชืออาจารย์ที่ปรึกษา
ปีการศึกษา 2540	ลายมือชืออาจารย์ที่ปรึกษาร่วม

C718241 : MAJOR COMPUTER SCIENCE
KEY WORD:
INDUCTIVE LOGIC PROGRAMMING / MACHINE LEARNING / PROGOL / RECOGNITION / THAI CHARACTERS

APINYA SUPANWASSA : AN APPLICATION OF INDUCTIVE LOGIC PROGRAMMING TO THAI PRINTED CHARACTER RECOGNITION. THESIS ADVISOR : DR. BOONSERM KIJSIRIKUL.

78 pp. ISBN 974-637-943-7.

The purpose of this thesis is to apply Inductive Logic Programming (ILP) to Thai printed character recognition. The ILP system which have been chosen is PROGOL. First, PROGOL is employed to learn Thai printed characters. The examples of characters and background knowledge are mainly used to train PROGOL. The output of PROGOL is a set of rules each of which defines the characteristics of a Thai character. Then, the learned rules are used to recognize an input character by comparing them with the input character, and the most matched rule together with associated character is selected as the output. Two experiments were run to test the method. In the first experiment designed for unseen fonts, the Eucrosia fonts are used for training and the Cordia fonts are used for testing. The recognition rate is 87.38%, tested with 539 characters. In the second experiment for noisy fonts, the Cordia and Eucrosia fonts are used for training, and the copies produced by a copy machine of both types of fonts composed of 2156 characters are used for testing. The recognition rate is 87.89%. The average recognition time is 0.13 second per character.

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต อภิญญา ศุภวรรณวรรณา
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่ออาจารย์ที่ปรึกษา
ปีการศึกษา 2540	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ อาจารย์ ดร.บุญเสริม กิจศิริกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่างๆ ของการวิจัยมาด้วยดีตลอด ขอขอบคุณ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล และ คุณเดชา รัตนาธาร ที่อนุญาติให้ใช้โปรแกรมการประมวลผลขั้นต้นจากงานวิทยานิพนธ์ของคุณเดชา รัตนาธาร ได้ ขอขอบคุณ คุณอดิศร ทิพยไพศาล ที่ให้ใช้เครื่องกวาดตรวจในการเก็บข้อมูลภาพตัวอักษร รวมทั้งคุณสรรชัย ธนาชัยเสง และเพื่อนๆที่คอยให้กำลังใจมาโดยตลอด

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งได้ให้การสนับสนุน และให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สารบัญ

หน้า

สารบัญ (ต่อ)

สารบัญตาราง

สารบัญภาพ

รูปที่ หน้า

สารบัญภาพ (ต่อ)