

**PRINTED THAI CHARACTER RECOGNITION  
BY GENETIC ALGORITHM**

**MONTRI DAVELOH**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY  
2005**

**ISBN 974-04-6586-2  
COPYRIGHT OF MAHIDOL UNIVERSITY**

การรู้จำตัวพิมพ์อักษรไทยด้วยหลักการทางพันธุศาสตร์ (PRINTED THAI CHARACTER RECOGNITION BY GENETIC ALGORITHM)

มนตรี เดวีเลาะ 4337416 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : ชมทิพ พรพนมชัย, Ph.D., สุกสงวน งามสุริยโรจน์, Ph.D.,  
สุกัญญา พงษ์สุภาพ, Ph.D.

บทคัดย่อ

งานวิจัยนี้ได้นำทฤษฎีทางพันธุกรรมในรูปแบบของ Cell Automata ผ่านกฎของ Conway's game of life มาใช้เพื่อสร้างระบบรู้จำตัวพิมพ์อักษรไทย โดยระบบที่เสนอนี้ประกอบด้วยสองส่วนหลัก คือ ส่วนของการฝึกการรู้จำ และส่วนของการทดสอบการรู้จำ ซึ่งภาพตัวอักษรพิมพ์ที่ป้อนให้กับส่วนแรกจะเป็นภาพที่ได้จากรูปแบบอักษรมาตรฐานในเครื่องคอมพิวเตอร์ที่เป็นที่นิยมใช้กันในปัจจุบันจำนวน 72,864 ตัวอักษร ส่วนรูปภาพที่นำมาใช้ทดสอบการรู้จำนั้นได้จากรูปภาพที่ได้จากการจับ (capture) หน้าจอภาพที่เก็บในรูปแบบ BMP จำนวน 1,015 ตัวอักษร

จากการศึกษาพบว่ารูปแบบที่ใช้ในการจัดเก็บเป็นฐานข้อมูลมีขนาดใหญ่โดยงานวิจัยนี้ได้แปลงค่าจากตาราง 64 X 64 มาจัดเก็บในรูปแบบ bit string ที่ทั้งนี้การใช้ตารางขนาด 64 X 64 เพื่อสร้างโอกาสการกระจายรูปแบบของ Generation สุดท้ายของแต่ละตัวอักษรให้มีความหลากหลายและเป็นเอกลักษณ์ของแต่ละตัวอักษรเด่นชัดขึ้น จึงทำให้กระบวนการในการปรับเปลี่ยนในแต่ละ Generation จนถึง Generation สุดท้ายใช้เวลานาน และฐานข้อมูลที่ใช้เป็นตัวแทนประชากรของ Generation สุดท้ายของแต่ละตัวต้องมีจำนวนมากพอที่จะทำให้รูปแบบของ bit string นั้นกลายเป็นรูปแบบเด่นของตัวอักษรเหล่านั้น ซึ่งจะทำให้ระบบสามารถทำการแยกแยะตัวอักษรนั้นได้จากจำนวนความถี่ของตัวอักษรที่เกิดรูปแบบ bit string นั้นมากที่สุด

ผลการทดสอบตัวอย่างจำนวน 1,015 ตัวอักษรสามารถได้ผลรู้จำถูกต้อง (Accept) จำนวน 986 ตัวอักษรคิดเป็น 97.14% ปฏิเสธ (Reject) จำนวน 6 ตัวคิดเป็น 0.59% และผิดพลาด (Miss) จำนวน 23 ตัวอักษรคิดเป็น 2.27% โดยมีความเร็วเฉลี่ยของการรู้จำประมาณ 85 วินาทีต่อตัวอักษร

85 หน้า ISBN 974-04-6586-2

**PRINTED THAI CHARACTER RECOGNITION BY GENETIC ALGORITHM****MONTRI DAVELOH 4337416 SCCS/M****M.Sc. (COMPUTER SCIENCE)****THESIS ADVISORS: CHOMTIP PORNPANOMCHAI, Ph.D., SUDSANGUAN NGAMSURIYAROJ , Ph.D., SUKANYA PHONGSUPHAP , Ph.D.****ABSTRACT**

This research applied a genetic algorithm in the pattern of cell automata and through Conway's rules of the game of life, to generate a system of printed Thai character recognition. The system consisted of two main parts, namely, recognition training and recognition testing. The printed character images fed to the first part were derived from standard character patterns widely used in a computer currently totalling 72,864 characters. As for the images used for recognition testing, they were captured from a computer screen and stored in BMP pattern, amounting to 1,015 characters.

The findings in this research revealed that the database used was of large size and data was transformed from a table frame of 64 x 64 pixels to be stored in the form of bit strings. A table size of 64 x 64 pixels was used to enable a wide variety of distribution patterns of the stable state of each character, making its identity more obvious. This, of course, caused a modification process in each generation till the final generation which took a long time while the database was used to represent the population of the final generation of each character must be large enough for the bit string used to represent these characters. This would enable the system to recognize a character based on its frequency with the largest number of those bit string patterns.

Out of 1,015 printed Thai characters tested, it was found that the system could recognize (accept) 986 characters or 97.14%, while rejecting 6 characters or 0.59% and misrecognizing 23 characters or 2.27%. The recognition speed is 85 seconds per character on the average.

**KEY WORDS: GENETIC ALGORITHM / CONWAY'S GAME OF LIFE /  
PRINTED THAI CHARACTER RECOGNITION****85 p. ISBN 974-04-6586-2**