

งานวิจัยในวิทยานิพนธ์ฉบับนี้เป็นการประยุกต์ใช้ทฤษฎีกราฟเซตแบบโทลเอนต์กับงานทางด้านความรู้จำตัวอักษรตัวพิมพ์ภาษาไทย ระบบที่นำเสนอนี้ยอมให้มีหลายกลุ่มย่อยต่อหนึ่งพยัญชนะตัวอักษร โดยแต่ละกลุ่มย่อยเทียบได้กับ Equivalent Class ในทฤษฎีกราฟเซต และจาก Equivalent Class นั้นสามารถแบ่งชนิดของ Class ออกได้เป็น 2 ชนิดคือ Lower Approximation Set และ Upper Approximation Set โดย Lower Approximation Set จะเป็นกลุ่มที่มีพยัญชนะตัวอักษรเดียวกันนั้นเป็นสมาชิก ส่วนกลุ่มที่มีสมาชิกตั้งแต่ 2 พยัญชนะตัวอักษรขึ้นไปจะถูกจัดเป็น Upper Approximation Set โดยในการสร้างกลุ่มย่อยหนึ่ง ๆ ใช้วิธีการวัดระยะห่างระหว่างแต่ละตัวอักษร โดยตัวอักษรที่ห่างกันน้อยกว่าหรือเท่ากับค่าที่กำหนดไว้จะถูกจัดในกลุ่มเดียวกัน ผลการทดสอบการรู้จำกับตัวอักษรตัวพิมพ์ภาษาไทยจำนวน 4 รูปแบบ (Font) ได้แก่ AngsanaUPC BowalliaUPC CordiaUPC และ FreesiaUPC โดยรูปแบบละ 4 ขนาดคือ 14 16 18 และ 22 พอยท์ (Point) ซึ่งจะได้ตัวอักษรรวมทั้งสิ้น 7040 ตัว ซึ่งจะแบ่งเป็นกลุ่มย่อย ๆ ได้ทั้งหมด 761 กลุ่มและในจำนวนนี้เป็น Upper Approximation Set ประมาณ 5% เมื่อนำระบบที่นำเสนอนี้เปรียบเทียบกับระบบการรู้จำแบบใช้โครงข่ายประสาทเทียมทั้ง 4 Configurations ได้แก่ โครงข่ายประสาทเทียมที่มี 36 input nodes/1 Hidden Layer/73 nodes ต่อ Hidden Layer/6 output nodes, 36 input nodes/1 Hidden Layer/73 nodes ต่อ Hidden Layer/44 output nodes, 36 input nodes/1 Hidden Layer/146 nodes ต่อ Hidden Layer/44 output nodes, 36 input nodes/2 Hidden Layer/73 nodes ต่อ Hidden Layer/44 output nodes ซึ่งให้ความถูกต้องในการรู้จำ 90.74%, 94.46%, 94.91% และ 95.04% ตามลำดับ สำหรับระบบการรู้จำแบบที่ใช้ Tolerant Rough Sets เป็นตัวรู้จำนั้นมีความถูกต้องในการรู้จำ 99.01% ซึ่งแสดงให้เห็นว่าระบบที่นำเสนอนี้มีประสิทธิภาพในการรู้จำสูงกว่าระบบแบบที่ใช้โครงข่ายประสาทเทียม

ABSTRACT

TE138920

In this thesis, we propose an application of Tolerant Rough Sets to recognition of Printed Thai Characters. The proposed system allows more than one group per character type. Each group is equal to an Equivalent Class in Rough Sets Theory. There are two kinds of classes, i.e., Upper Approximation Set and Lower Approximation Set. In Lower Approximation Sets, there exists only one type of characters. On the other hand, Upper Approximation Sets contain more than one type of character as their members. To measure the distance between each character in the Equivalent Class. Characters with the measured distance less than the predefined threshold value are grouped together. The characters used for experiment in this thesis consist of 4 fonts, AngsanaUPC, BrowalliaUPC, CordiaUPC, and FreesiaUPC, each of which has 4 sizes, 14, 16, 18 and 22 points, total of 7040 characters. From this characters 761 groups are formed and less than 4% of these groups are Upper Approximation Sets. In this work, the proposed system is compared to the Artificial Neural Networks system. Four configurations of the Artificial Neural Networks are used in this work, which give the recognition rate of 90.74%, 94.46%, 94.91% and 95.04%. For the proposed system the recognition accuracy is 99.01%. This can be considered that the proposed system has more advantages over the Artificial Neural Networks system.