หัวข้อวิทยานิพนธ์	อัลกอริทึมใหม่สำหรับการจำแนกประเภทเอกสาร โดยใช้กฎความสัมพันธ์
นักศึกษา	นางสาว สุภาภรณ์ บุตรดีวงษ์
รหัสนักศึกษา	45066003
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
พ.ศ.	2549
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนออัลกอริทึมใหม่สำหรับใช้จำแนกเอกสารโดยใช้กฎความสัมพันธ์ (Association Rule-Based Text Classifier: ARTC) มีวัตถุประสงค์เพื่อปรับปรุงการจำแนกเอกสารของอัลกอริทึม Association Rule-based Classifier By Categories (ARC-BC) ซึ่งเป็นการที่ใช้กฎความสัมพันธ์ในการจำแนกประเภทเอกสารโดยใช้กฎความสัมพันธ์ที่สามารถจำแนกเอกสารได้ดีหากเอกสารนั้น ๆ มีลักษณะเอกสารที่กลุ่มไม่ซ้อนทับกัน แต่ในความเป็นจริงของเอกสารที่มีอยู่ในปัจจุบันนั้นมีเอกสารทั้งที่ไม่มีกลุ่มซ้อนทับกัน และกลุ่มซ้อนทับกัน ดังนั้นวิทยานิพนธ์ฉบับนี้จึงได้นำเสนอวิธีการจำแนกเอกสารโดยใช้กฎความสัมพันธ์ที่สามารถจำแนกเอกสารที่มีลักษณะซ้อนทับกันและไม่ซ้อนทับกันได้เป็นอย่างดี โดยเริ่มจากการค้นหากฎความสัมพันธ์ของข้อมูล ซึ่งได้ Frequent Itemset 2 ชนิด คือ Frequent Itemset ที่เก็บคุณลักษณะของเอกสาร ที่ไม่ซ้อนทับกับกลุ่มใด ๆ และ Frequent Itemset ที่เก็บคุณลักษณะของเอกสารที่ซ้อนทับกันมากกว่าหนึ่งกลุ่ม ซึ่งเกิดจากการเชื่อมความสัมพันธ์แบบใหม่ที่เรียกว่า ARTC join เมื่อได้ความสัมพันธ์ที่จะนำไปสร้างเป็นกฎความสัมพันธ์แล้ว นำกฎความสัมพันธ์ที่ได้มาคัดทิ้ง โดยใช้โครงสร้างต้นไม้ ในการคัดกฎความสัมพันธ์ที่ไม่จำเป็นทิ้ง ก่อนที่จะนำกฎที่ได้ไปจำแนกประเภทเอกสาร ผลการทดลองที่ได้อัลกอริทึมที่นำเสนอ ให้อัตราความถูกต้อง (Accuracy rate)ในการจำแนกประเภทเอกสารได้ดีกว่าอัลกอรึทึม ARC-BC

Thesis Title	Text Categorization using a new Association Rule-Based Classifier Algorithm
Student	Miss Supaporn Buddeewong
Student ID.	45066003
Degree	Master of Science
Programme	Information Technology
Year	2006
Thesis Advisor	Asst.Prof.Dr.Worapoj Kreesuradej

ABSTRACT

This thesis proposes a new Association Rule-Based Text Classifier (ARTC) algorithm to improve the prediction accuracy of Association Rule-based Classifier By Categories (ARC-BC) algorithm. ARC-BC has shown a good performance. In addition, the classifier based on ARC-BC algorithm produces clear and understandable results. However, the classifier can not work well for the single-class document that has some terms of document mutually associated with other classes. Unlike ARC-BC algorithm, a new Association Rule-Based Text Classifier (ARTC) algorithm consists of three main phase to construct a classifier. The first phase is association rule generation. The proposed association rule generation algorithm constructs two types of frequent itemsets. The first frequent itemset contain all terms that have no an overlap with other categories. The second frequent itemset contain all features that have an overlap with other categories that generated by a new join method, ARTC join. The second pahse is pruning step. The pruning step uses tree structure for pruning association rule that have confidence value more that a threshold value of confidence factor. The last phase is the prediction of classes associated with new documents. The experimental results are shown a good performance of the proposed classifier.