

วิทยานิพนธ์นี้เสนอการทดสอบประสิทธิภาพของระบบการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ร่วมกับกฎความสัมพันธ์และผลสะท้อนกลับจากผู้ใช้ โดยจะเปรียบเทียบกับระบบค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์และระบบการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ร่วมกับกฎความสัมพันธ์ของคำ ซึ่งในเทคนิคปริภูมิเวกเตอร์จะมีการแปลงเอกสารและข้อสอบถามให้อยู่ในรูปของเวกเตอร์ ส่วนเทคนิคกฎความสัมพันธ์เป็นเทคนิคของการทำเหมืองข้อมูล โดยหาความสัมพันธ์ของคำที่เกิดขึ้นพร้อมกันบ่อยครั้งในเอกสาร เพื่อเพิ่มคำที่มีความสัมพันธ์กับคำในข้อสอบถามก่อนนำไปใช้ดึงเอกสาร ส่วนเทคนิคผลสะท้อนกลับจากผู้ใช้คือเทคนิคที่ใช้ผลสะท้อนกลับจากผู้ใช้ในการปรับข้อสอบถามให้เข้าใกล้กลุ่มเอกสารที่เกี่ยวข้องกับข้อสอบถามมากยิ่งขึ้น

งานวิจัยนี้เป็นงานวิจัยเชิงทดลอง โดยใช้เอกสารนิตยสาร TIME จำนวน 425 เอกสารและข้อสอบถามจำนวน 83 ข้อสอบถามทดลองเปรียบเทียบประสิทธิภาพของระบบค้นคืนเอกสารโดยการคำนวณค่าเฉลี่ยฮาร์โมนิคของระบบค้นคืนเอกสารทั้ง 3 รูปแบบ ดังกล่าวข้างต้น

จากการวิเคราะห์ผลการทดลองสรุปได้ว่าระบบค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ร่วมกับกฎความสัมพันธ์ของคำสามารถทำให้ประสิทธิภาพดีขึ้นกว่าการใช้เทคนิคปริภูมิเวกเตอร์ แต่เมื่อใช้เทคนิคปริภูมิเวกเตอร์ร่วมกับกฎความสัมพันธ์และผลสะท้อนกลับจากผู้ใช้ทำให้ประสิทธิภาพของระบบการค้นคืนเอกสารมากกว่าการใช้เทคนิคปริภูมิเวกเตอร์เพียงอย่างเดียว แต่ต่ำกว่าการใช้เทคนิคปริภูมิเวกเตอร์ร่วมกับเทคนิคการใช้กฎความสัมพันธ์ของคำ

This thesis presents an experimental study on using an information retrieval system that employs a vector space technique together with association rules and relevant feedback in comparison with a system that uses the vector space technique alone and a system that uses the vector space technique together with association rules. In vector space technique, documents and queries are transformed to be vectors, while association rules is a data mining technique that is used to find associations of words that appear in the same documents. The list of associated words can be used to extend the query vector before using it to retrieve a set of relevant documents. Such query vector can be refined further by applying relevant feedback which is a technique that adjusts the query vector according to user feedback on the list of documents which is the result from the first round of query. This is to make the query vector closer to the target documents.

In the thesis, the performance of the three information retrieval systems above is compared through Harmonic mean. The experiments were conducted on 425 documents and 83 queries from the TIME Magazine collection which is obtained from <ftp://ftp.cs.cornell.edu/pub/smart/time>.

The experimental results show that the information retrieval system that uses vector space with association rules has a best performance while the one using vector space together with association rule and relevant feedback shows a better performance than the one using vector space alone.