

บทคัดย่อ

ระบบการค้นคืนสารสนเทศที่ใช้คำในข้อความถาม (Query) ไปจับคู่กับคำหลัก (Keyword) ในเอกสารอาจไม่มีประสิทธิภาพเพียงพอ เพราะคำหลักไม่สามารถเป็นตัวแทนเอกสารได้อย่างครบถ้วน ผลของการค้นคืนจึงจะได้เอกสารที่มีคำหลักตรงกับคำในข้อความถามของผู้ใช้เท่านั้น แต่ไม่ใช่เป็นการค้นคืนในเชิงความหมายที่มีอยู่ในเอกสาร อีกวิธีการหนึ่งคือการใช้เทคนิคเชิงแมทริกซ์ที่มีการเปลี่ยนรูปแบบการเก็บคำสำคัญของเอกสารในรูปแบบแมทริกซ์ของคำสำคัญสำหรับเอกสาร และใช้เทคนิคการกระจายแมทริกซ์เพื่อให้ได้กลุ่มของคำหลักที่มีความสัมพันธ์กับเอกสาร งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบกระจายแมทริกซ์สองแบบในการค้นคืนสารสนเทศข้อความภาษาไทย อันได้แก่ การกระจายแมทริกซ์แบบเซมิดีสครีต (Semidiscrete Matrix Decomposition) และการแยกค่าแบบเดี่ยว (Singular Value Decomposition) สำหรับในเอกสาร 4 สาขาวิชา ผลการทดลองพบว่า การกระจายแมทริกซ์ด้วยวิธีเซมิดีสครีตใช้เนื้อที่ในการกระจายแมทริกซ์ (Decomposition storage) น้อยกว่าการกระจายแมทริกซ์ด้วยวิธีแยกค่าแบบเดี่ยวอย่างมีนัยสำคัญ แต่ใช้เวลาในการกระจายแมทริกซ์ (Decomposition time) มากกว่า ส่วนผลการสืบค้นสารสนเทศอันได้แก่ ความแม่นยำ (Precision) และ การเรียกกลับ (Recall) ของการกระจายแมทริกซ์ด้วยวิธีเซมิดีสครีตและด้วยวิธีแยกค่าแบบเดี่ยวจะขึ้นอยู่กับารปรับแต่งองค์ประกอบต่าง ๆ ในกระบวนการออกแบบและพัฒนาค้นคืนสารสนเทศ