

ชื่อวิทยานิพนธ์ : การออกแบบดรรชนีผกผันที่สามารถรองรับการเพิ่มขึ้นของเอกสาร
ชื่อผู้เขียน : น.ส. ศิริภิญโญ จันทมณี
ชื่อปริญญา : วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
ปีการศึกษา : 2546

ระบบการสืบค้นสารสนเทศทำงานในสิ่งแวดล้อมที่มีการเปลี่ยนแปลงบ่อยครั้ง ซึ่งมีการเพิ่ม ลบ และแก้ไขเอกสาร ระบบสืบค้นสารสนเทศแบบดั้งเดิมจะทำการสร้างดรรชนีของเอกสารที่มีอยู่ทั้งหมดใหม่ในทุกครั้งที่มีการปรับปรุงข้อมูล ซึ่งจะเป็นวิธีที่มีประสิทธิภาพต่ำสำหรับระบบที่มีเอกสารจำนวนมาก จึงมีการเสนอแนวคิดของการปรับปรุงดรรชนีเพียงส่วนที่มีการเปลี่ยนแปลงเท่านั้น การปรับปรุงข้อมูลของระบบที่มีปัญหามากที่สุด คือ การเพิ่มเอกสารใหม่เข้ามาในระบบ เนื่องจากต้องมีการขยายและปรับเปลี่ยนดรรชนีที่กำลังใช้งานอยู่ ทำให้ไม่สามารถค้นหาเอกสารได้ในขณะที่มีการปรับปรุงข้อมูล

งานวิจัยนี้มุ่งประเด็นการปรับปรุงข้อมูลเมื่อมีการเพิ่มเอกสารใหม่เข้ามาในระบบ โดยที่การจัดสรรเนื้อที่เพื่อรองรับการขยายตัวนั้นจะพิจารณาจากพารามิเตอร์ 2 ค่า คือ ความสม่ำเสมอในการพบเจอ และความถี่โดยเฉลี่ยของเจอ เสนอแนวคิดในการจัดสรรเนื้อที่ในรูปแบบของแผนการจัดสรรเนื้อที่จำนวน 4 แผน แต่ละแผนจะมีลักษณะการจัดสรรเนื้อที่ที่แตกต่างกัน และเสนอวิธีการจัดการเนื้อที่เพื่อให้การสืบค้นเป็นไปอย่างมีประสิทธิภาพ

Information retrieval systems operate in dynamic environments where documents in the systems can be deleted, modified and inserted into the systems. The traditional approach of completely rebuilding the entire inverted indices may not be feasible especially with large systems. Therefore, the indices have to support in-place updates. Among the three operations, the most problematic one is the incremental updates since it requires the expansion and readjustment of the index currently in use. This may disrupt the availability of the index during the updating process.

In this research, we focus on incremental update operation. The amount of space allocated to accommodate expansion is based upon to two parameters: the consistency of term appearance and the average term frequency. We propose four space allocation plans. Each plan has a distinct set of allocation functions based on its characterization. Furthermore, space management strategy are purposed to improve the retrieval performance.