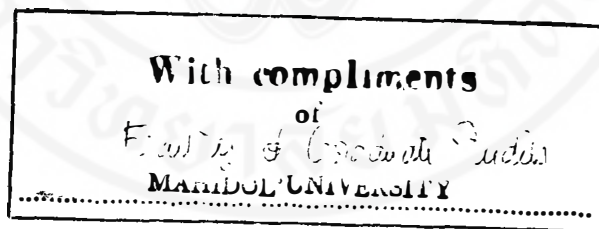


610855721

**USING SIGNATURE FILE TECHNIQUE FOR SEARCHING  
A MULTI-VALUED ATTRIBUTE IN  
RELATIONAL DATABASE**

**PAIRUT LEELAHAKARNJANA**



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY**

**1998**

**ISBN 974-661-055-4**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

TF  
E1589  
1009

Copyright by Mahidol University

SCCS/M3937001 : MAJOR : COMPUTER SCIENCE ; M.Sc. (COMPUTER SCIENCE)

KEYWORD : SIGNATURE FILE / MULTI-VALUED ATTRIBUTE /  
RELATIONAL DATABASE

PAIRUT LEELAHAKARNJANA : USING SIGNATURE FILE TECHNIQUE FOR  
SEARCHING A MULTI-VALUED ATTRIBUTE IN RELATIONAL DATABASE. THESIS  
ADVISOR : DAMRAS WONGSAWANG Ph.D. 78 p. ISBN 974-661-055-4

Signature file is an indexing technique for searching unformatted data in an information retrieval system. The earliest structure of signature file is sequential signature file. In query evaluation, sequential signature file requires accessing all bits in the signature file. Many researches proposed various organizations for the signature file that try to access without all bits. One interesting organization is vertically partitioned signature file. Recently, the multi-fragmented signature file is proposed for fast searching in the multi-term query. It is also grouped in the vertically partitioned signature.

The multi-valued attribute in relational database can be viewed as unformatted data, and querying the multi-valued attribute by using SQL (Structural Query Language) requires high processing time. For these reasons, we propose a model that uses signature file for searching the multi-valued attribute in a relational database. Generally, if a relation has the multi-valued attribute, it should be split into many relations for getting normal forms. In our proposed model, the multi-valued attribute can be kept in a single relation. The signature file is created to be the index of the multi-valued attribute. We simulate the test environments and compare the results between our approach with the SQL statement approach that use SQL statement for getting the same result. We found that our approach gives faster response time than the SQL statement approach. The reason behind this result is because signature file processing deals with bit operation which always takes less processing time than record operation used in SQL.

For SQL statement approach, the factors that affect the query response time are database size and number of terms in the query. While in our approach, the query response time depends on database size, number of projects in query, and number of records in the output. Increasing database size usually increases the query response time in both approaches. Increasing the number of terms in query increases the query response time in SQL statement approach, but it decreases the query response time in our approach. The reason is because increasing the number of terms in query always reduces the number of records in output and the increased time when increasing number of terms in query is always less than the decreased time when decreasing number of records in output. Therefore, for our approach, the query response time is always decreased when the number of terms in a query is increased.

SCCS/M3937001 : สาขาวิชา : วิทยาการคอมพิวเตอร์; วท.ม (วิทยาการคอมพิวเตอร์)

ไพรัตน์ ลิพหากาญญา : การใช้เทคนิคของแฟ้มข้อมูลลายมือในการค้นหาแบบพหุเทอมในฐานข้อมูลสัมพันธ์ (Using signature file for searching a multi-valued attribute in relational database) คณะกรรมการควบคุมวิทยานิพนธ์ : คำรัส วงศ์สว่าง Ph.D. , ศุภชัย ตั้งวงศ์ศานย์ Ph.D., จุฬารัตน์ ตันประเสริฐ Ph.D. 78 หน้า. ISBN 974-661-055-4

แฟ้มข้อมูลลายมือเป็นวิธีการหนึ่งที่ใช้ในการทำดัชนีของข้อมูลที่ไม่มีรูปแบบ โครงสร้างแรกที่ถูกเสนอจะประกอบด้วยข้อมูลลายมือเรียงกันไปอย่างต่อเนื่อง ในการค้นหาแบบพหุเทอมทุกบิตของแฟ้มข้อมูลลายมือจะถูกตรวจสอบ งานวิจัยจำนวนมากได้ทำการศึกษาและพัฒนาโครงสร้างของแฟ้มข้อมูลลายมือโดยพยายามลดจำนวนบิตที่ต้องการตรวจสอบ โครงสร้างหนึ่งที่น่าสนใจคือการแบ่งส่วนของแฟ้มข้อมูลลายมือตามขวาง

เนื่องจากแบบพหุเทอมในฐานข้อมูลสัมพันธ์สามารถถูกจัดเป็นข้อมูลที่ไม่มีรูปแบบ และการใช้ SQL ในการค้นหาแบบพหุเทอมต้องการเวลาในการประมวลผลมาก ดังนั้น งานวิทยานิพนธ์นี้จึงนำเสนอการใช้แฟ้มข้อมูลลายมือในการทำดัชนีเพื่อใช้ในการค้นหาแบบพหุเทอมในฐานข้อมูลสัมพันธ์ ในการออกแบบฐานข้อมูลสัมพันธ์ ถ้าฐานข้อมูลมีแบบพหุเทอม แบบพหุเทอมจะถูกจัดเก็บแยกต่างหาก แต่ในวิธีการที่นำเสนอนี้ แบบพหุเทอมสามารถถูกจัดเก็บรวมในฐานข้อมูลเดิมได้ และแฟ้มข้อมูลลายมือจะถูกสร้างขึ้นเพื่อใช้เป็นดัชนีของแบบพหุเทอม จากการทดลองพบว่า วิธีการใช้แฟ้มข้อมูลลายมือใช้เวลาในการประมวลผลน้อยกว่าวิธีการใช้ SQL สาเหตุที่ทำให้เกิดผลดังกล่าว คือ วิธีการใช้แฟ้มข้อมูลลายมือจะเกี่ยวข้องกับการจัดการกับบิตในแฟ้มข้อมูลลายมือ ในขณะที่วิธีการใช้ SQL จะเกี่ยวข้องกับการจัดการกับเรคคอร์ดของฐานข้อมูล และเวลาที่ใช้ในการจัดการกับบิตจะน้อยกว่าเวลาที่ใช้ในการจัดการกับเรคคอร์ดของฐานข้อมูล

ปัจจัยที่มีผลต่อเวลาที่ใช้ในการค้นหาแบบพหุเทอมของวิธีการใช้ SQL คือขนาดของฐานข้อมูล จำนวนเทอมที่ค้นหา ส่วนของวิธีการใช้แฟ้มข้อมูลลายมือขึ้นกับขนาดของฐานข้อมูล จำนวนเทอมที่ค้นหา และจำนวนเรคคอร์ดในผลลัพธ์ การเพิ่มขนาดของฐานข้อมูลจะทำให้เวลาในการค้นหาเพิ่มขึ้นในทั้งสองวิธี การเพิ่มจำนวนเทอมที่ค้นหาจะเพิ่มเวลาในการค้นหาของวิธีการใช้ SQL แต่จะลดเวลาในการค้นหาของวิธีการใช้แฟ้มข้อมูลลายมือ เนื่องจากการเพิ่มจำนวนเทอมที่ค้นหาทำให้จำนวนเรคคอร์ดในผลลัพธ์ลดลงเสมอ และเวลาที่เพิ่มขึ้นเนื่องจากการเพิ่มจำนวนเทอมที่ค้นหามักน้อยกว่าเวลาที่ลดลงเนื่องจากการลดลงของจำนวนเรคคอร์ดในผลลัพธ์