การเก็บและการค้นคืนสารสนเทศโดยใช้แนวความคิดของแฟ้มข้อมูลผกผัน

นางสาว พรทิพย์ บัวสาม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2535

ISBN 974-581-197-1

INFORMATION STORAGE AND RETRIEVAL USING INVERTED FILE

MISS PORNTIP BUASAM

A Thesis Submitted in Patial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Computer Engineering
Graduate School
Chulalongkorn University
1992

ISBN 974-581-197-1

หัวข้อวิทยานิพนธ์	การเก็บและการค้นคืนสารสนเทศโดยใช้แนวความคิดของแฟ้มข้อมูลผกผัน
โดย	นางสาว พรทิพย์ บัวสาม
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ จารุมาตร ปิ่นทอง

บัณฑิตวิทยาลัยจุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ เดือน สินธุพันธ์ประทุม)

..... อาจารย์ที่ปรึกษา
(อาจารย์ จารุมาตร ปิ่นทอง)

..... กรรมการ
(รองศาสตราจารย์ มัณฑนา ปรากาศสมุทร)

..... กรรมการ
(อาจารย์ ดร.ธรรยง เต็งอำนวย)

พรทิพย์ บัวสาม : การเก็บและการค้นคืนสารสนเทศโดยใช้แนวความคิดของแฟ้มข้อมูลผกผัน (INFORMATION STORAGE AND RETRIEVAL USING INVERTED FILE) อาจารย์ที่ปรึกษา : อาจารย์ จารุมาตร ปิ่นทอง, 77 หน้า ISBN 974-581-197-1

วัตถุประสงค์ของวิทยานิพนธ์นี้ เพื่อศึกษาวิธีการเก็บ และการค้นคืนสารสนเทศที่เป็นข้อความ โดยอาศัยแนวความคิดของแฟ้มข้อมูลผกผันเป็นพื้นฐาน และพัฒนาโปรแกรมคอมพิวเตอร์โดยใช้ภาษาซี การออกแบบการทำงานของโปรแกรม แบ่งออกเป็น 2 ส่วน คือ ส่วนของการสร้างฐานข้อมูล และส่วนของการค้นคืนฐานข้อมูล ส่วนของการสร้างฐานข้อมูล จะทำการอ่านข้อความในแฟ้มข้อมูลข้อความที่ผ่านการแบ่งคำแล้ว โดยที่แฟ้มข้อมูลข้อความสามารถสร้างโดยโปรแกรมเวิร์ดโปรเซสเซอร์ทั่วไป และทำการสร้างแฟ้มข้อมูลดิคชันนารี เพื่อเก็บคำที่เป็นดัชนี และสร้างแฟ้มข้อมูลผกผัน เพื่อเก็บรายการหมายเลขอ้างอิงตำแหน่งของเอกสารที่สัมพันธ์กับดัชนี และส่วนของการค้นคืนฐานข้อมูล จะเป็นการค้นคืนเอกสารที่ต้องการ โดยในการค้นคืน เอกสารที่ต้องการจะถูกระบุโดยดัชนี ดัชนีจะถูกค้นหาในแฟ้มข้อมูลดิคชันนารี และเอกสารในแฟ้มข้อมูลข้อความ ที่ถูกระบุโดยหมายเลขอ้างอิงที่สัมพันธ์กับดัชนี ก็จะถูกอ่านออกมาจากแฟ้มข้อมูลข้อความ ในการออกแบบแฟ้มข้อมูล จะใช้โครงสร้างข้อมูลแบบบีทรี แทนแฟ้มข้อมูลดิคชันนารี และใช้โครงสร้างข้อมูลแบบลิงค์ลิส แทนแฟ้มข้อมูลผกผัน ผลที่ได้จากการทำงานของโปรแกรมนี้ ผู้ใช้สามารถค้นคืนเอกสารที่ต้องการได้สะดวก รวดเร็ว และถูกต้อง

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา 2534 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

C116893 : MAJOR COMPUTER SCIENCE
KEY WORD : INFORMATION RETRIEVAL/INVERTED FILE
PORNTIP BUASAM : INFORMATION STORAGE AND RETRIEVAL USING INVERTED FILE.
THESIS ADVISOR : CHARUMATR PINTHONG, 77 pp. ISBN 974-581-197-1

The purpose of this thesis is to study text information storage and retrieval based on inverted file concept, and to develope computer program using C language compiler. The program consists of two subsystem : the database creation subsystem, and the database retrieval subsystem. Database creation subsystem scans a word seperated text file, created by using any word processor, and then creates a dictionary file for storing the allowable indexing terms, and creates a inverted file for storing an associated list of document reference numbers of each index. Database retrieval sumsystem retrieves documents from the stored text file. Required documents are identified by an arbitrary index, index requires a dictionary search to find the associated document reference numbers, and the identified documents are selected from the text file. The dictionary file is implemented by using B-tree and the inverted file using linked list. This program can retrieve stored documents conveniently, rapidly, and accurate.

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา 2534 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ ได้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่ง ของอาจารย์ จารุมาตร ปิ่นทอง อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่างๆ ของการวิจัยมาด้วยดีตลอด

ขอขอบคุณเจ้าหน้าที่ บริษัทซัมมิตคอมพิวเตอร์ ที่ได้เอื้อเฟื้อให้ใช้โปรแกรมสำหรับตัดคำเอกสารภาษาไทย และคำแนะนำที่เป็นประโยชน์ต่อการวิจัยในครั้งนี้

เนื่องจากทุนการวิจัยครั้งนี้บางส่วนได้รับมาจาก ทุนอุดหนุนการวิจัยของบัณฑิตวิทยาลัย จึงขอขอบพระคุณบัณฑิตวิทยาลัยมา ณ ที่นี้ด้วย

ขอขอบคุณ พี่ๆ น้องๆ และเพื่อนๆ ที่ช่วยเหลือ และให้กำลังใจแก่ผู้วิจัยตลอดมา

ท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ซึ่งสนับสนุนและให้กำลังใจแก่ผู้วิจัยเสมอมา จนสำเร็จการศึกษา

สารบัญ

หน้า

หน้า

สารบัญภาพ

หน้า