# การค้นคืนสารสนเทศโดยใช้แนวคิดแบบจำลองปริภูมิเวกเตอร์

นาย สาโรช เมาลานนท์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2535

ISBN 974-579-986-6

# INFORMATION RETRIEVAL USING VECTOR SPACE MODEL CONCEPT

MR. SAROCH MOULANONT

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Computer Engineering

Graduate School

Chulalongkorn University

1992

ISBN 974-579-986-6

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | การค้นคืนสารสนเทศโดยใช้แนวคิดแบบจำลองปริภูมิเวกเตอร์ |
| โดย | นาย สาโรช เมาลานนท์ |
| ภาควิชา | วิศวกรรมคอมพิวเตอร์ |
| อาจารย์ที่ปรึกษา | อาจารย์ จารุมาตร ปิ่นทอง |

---

บัญฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัญฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
( ผู้ช่วยศาสตราจารย์ วิชาญ เลิศวิภาตระกูล )

..... อาจารย์ที่ปรึกษา
( อาจารย์ จารุมาตร ปิ่นทอง )

..... กรรมการ
( ผู้ช่วยศาสตราจารย์ สุเมธ วัชระชัยสุรพล )

..... กรรมการ
( ผู้ช่วยศาสตราจารย์ วันพร ปิ้นเก่า )

สาโรช เมาลานนท์ : การค้นคืนสารสนเทศโดยใช้แนวคิดแบบจำลองปริภูมิเวกเตอร์
(INFORMATION RETRIEVAL USING VECTOR SPACE MODEL CONCEPT)
อ.ที่ปรึกษา อาจารย์จารุมาตร ปิ่นทอง. 89 หน้า. ISBN 974-579-986-6

จุดประสงค์ของวิทยานิพนธ์ฉบับนี้ เพื่อศึกษากลไกที่ใช้ในการค้นคืนสารสนเทศ โดยอาศัยแนวคิดแบบจำลองปริภูมิเวกเตอร์เป็นพื้นฐาน พร้อมได้สร้างโปรแกรมเพื่อใช้ในการทดสอบประเมินผล ในวิทยานิพนธ์ฉบับนี้ได้กล่าวถึงความเป็นมาของการวิจัย ทฤษฎีที่เกี่ยวข้องกับการวิจัย ได้แก่ ความเป็นมาของระบบการจัดการข้อความ ระบบฐานข้อมูลข้อความ การประยุกต์ทางการจัดการข้อความ กระบวนการค้นคืนสารสนเทศทั้งแบบสัญนิยม และระบบการค้นคืนชั้นสูง การพัฒนาโปรแกรมโดยใช้ภาษาซีเป็นเครื่องมือ การออกแบบโปรแกรมได้คำนึงถึงส่วนประกอบพื้นฐานของระบบฐานข้อมูล คือ แบบจำลองข้อมูล บทนิยามข้อมูล และกลไกการค้นคืนสารสนเทศ ผลการทดสอบโปรแกรมปรากฏว่า การคำนวณค่าความใกล้เคียงระหว่างเวกเตอร์ของเอกสารกับประโยคสอบถามถูกต้องตรงตามผลที่คำนวณได้ด้วยมือ และสามารถค้นคืนเอกสารออกมาได้ตามลำดับค่าความใกล้เคียงจากมากไปหาน้อย ส่วนผลการทดสอบระบบการค้นคืนสารสนเทศโดยรวมพบว่า การทำดรรชนีเพื่อเป็นตัวแทนของเอกสาร มีความสำคัญต่อประสิทธิภาพของระบบการค้นคืนสารสนเทศอย่างมาก

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

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

## C017067 : MAJOR COMPUTER SCIENCE
KEY WORD : INFORMATION RETRIEVAL/VECTOR SPACE MODEL
SAROCH MOULANONT : INFORMATION RETRIEVAL USING VECTOR SPACE MODEL CONCEPT. THESIS ADVISOR : CHARUMATR PINTHONG. 89 pp.
ISBN 974-579-986-6

The objective of this thesis is to study the machanism of information retrieval based on vector space model concept and develop computer program, bibliographic database, for experimental evaluation. This thesis composes of the background research, theory related to the researches such as background of text management, text database, application of text management, conventional and advance information retrieval. The computer programs are written in C language. The design considers the basic components of database system that are data model, data definition and retrieval machanism. The result of program testing can be summarized as follows :- the computational similarity between the vector of documents and query compare with manual calculation are correct. The documents can be arranged by descending order of corresponding similarity with the query. The total results of the information retrieval system show that the indexing of documents representation is very significant for the efficiency of retrieval system.

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

# กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษาวิทยานิพนธ์ อาจารย์จารุมาตร ปิ่นทอง ที่ได้ให้คำปรึกษาแนะนำแนวทางที่เป็นประโยชน์ต่อการวิจัย พร้อมทั้งคอยผลักดันให้วิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วงด้วยดีมาตลอด ขอขอบพระคุณท่านอาจารย์ที่ปรึกษานิสิต รศ.สมชาย ทยานยง ที่ให้ความกรุณา ติดตามสอบถามความคืบหน้าของวิทยานิพนธ์อยู่เสมอมา และขอขอบพระคุณท่านอาจารย์ทุกท่านในภาควิชาวิศวกรรมคอมพิวเตอร์ ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้

นอกจากนี้ขอขอบคุณ คุณศุภกร โสภณวสุ ที่ได้ช่วยเหลือจัดเตรียมแผ่นใส และให้กำลังใจในระหว่างการทำวิทยานิพนธ์ ขอขอบคุณบรรดาเพื่อน ๆ พี่ ๆ น้อง ๆ ที่คอยให้ความช่วยเหลือ และกำลังใจต่อผู้วิจัยเสมอมา

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ซึ่งเป็นผู้มีพระคุณสูงสุดอันหาที่เปรียบมิได้ที่ได้ให้การอุปการะและกำลังใจต่อผู้วิจัยตลอดมา

สารบาญ

สารบาญ (ต่อ)

สารบาญ (ต่อ)

# สารบาญรูปประกอบ

สารบาญรูปประกอบ (ต่อ)