LIBRARY AND INFORMATION CENTER
MAHIDOL UNIVERSITY

OBJECT-ORIENTED MODELING APPROACH TO WORLD WIDE WEB SEARCHING

PAGAPORN EKAREEJIT

อภินันทนาการ
จาก
.....

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

(COMPUTER SCIENCE)

IN

FACULTY OF GRADUATE STUDIES

MAHIDOL UNIVERSITY

1996

Copyright© by Mahidol University

TH
P131e
1996

ชื่อวิทยานิพนธ์ การประยุกต์ใช้แบบจำลองโครงสร้างเชิงวัตถุเพื่อค้นหาข้อมูลในระบบ World Wide Web

ผู้วิจัย ผกาพร เอกอารีจิตร

ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์

ดำรัส วงศ์สว่าง Ph D.

ศุภชัย ตั้งวงศ์ศานต์ Ph D

วันที่สำเร็จการศึกษา 19 ธันวาคม พ.ศ 2539

บทคัดย่อ

ปัจจุบัน การค้นหาข้อมูลในระบบ World Wide Web ให้ตรงความต้องการเป็นสิ่งที่เป็นไปได้ยาก เนื่องจากมีจำนวนของข้อมูลมากมาย นอกจากนั้นโครงสร้างของการนำเสนอข้อมูลใน World Wide Web อยู่ในรูปแบบของ Hypertext และประกอบด้วยข้อมูล Multimedia ได้แก่ ข้อความ รูปภาพ ภาพเคลื่อนไหว และเสียง ดังนั้นการค้นหาข้อมูลผู้ใช้จึงต้องการโปรแกรมเพื่อช่วยการค้นหา

MUSEARCH คือโปรแกรมเพื่อช่วยการค้นหาข้อมูลให้ได้ข้อมูลตรงกับความต้องการ ประกอบด้วยข้อมูลที่อยู่ในรูปแบบ Mulitmedia ได้ โดยใช้หลักการของการประยุกต์ใช้แบบจำลองโครงสร้างเชิงวัตถุ (Object-Oriented Modeling) การประยุกต์ใช้ความสัมพันธ์แนวโครงสร้างของเทคนิคแบบจำลองโครงสร้างเชิงวัตถุ 4 หลักการได้แก่ Classification, Generalization, Aggregation และ Association สามารถเสนอรูปแบบข้อมูล 7 ชนิด ได้แก่ TEXT, IMAGE, AUDIO, VIDEO, APPLICATION, JAVAAPPLET และ JAVASCRIPT

โปรแกรมประกอบด้วยการทำงาน 4 ส่วน ได้แก่ ส่วนการติดต่อกับผู้ใช้ ส่วนการเก็บข้อมูลเข้าระบบฐานข้อมูล ส่วนการค้นหาข้อมูล และระบบฐานข้อมูล ส่วนการติดต่อกับผู้ใช้ ทำหน้าที่ติดต่อระหว่างผู้ใช้กับส่วนการเก็บข้อมูลเข้าระบบฐานข้อมูล และระหว่างผู้ใช้กับส่วนการค้นหาข้อมูล สำหรับส่วนการเก็บข้อมูลเข้าระบบฐานข้อมูล เมื่อเจ้าของแฟ้มข้อมูล HTML ลงทะเบียน ระบบจะสร้างข้อมูลประกอบด้วยประเภทข้อมูล และรายละเอียดของข้อมูลที่จำเป็นต่อการค้นหาข้อมูล ที่ประกอบอยู่ภายในแฟ้มข้อมูล HTML ส่วนการค้นหาข้อมูล รับการสอบถามจากผู้ใช้โดยสามารถสอบถามด้วย คำ สาขา แหล่งข้อมูล (URL) หรือประเภทข้อมูล และส่งผลการสอบถามกลับคืนผู้ใช้ ระบบฐานข้อมูลเก็บข้อมูล เก็บข้อมูลที่ได้จากส่วนการเก็บข้อมูล โปรแกรมใช้การจัดเรียงข้อมูลเรียกว่า Inverted Object Vectors เพื่อช่วยเพิ่มความเร็วในการค้นหา

การทดสอบ ด้วยตัวอย่างข้อมูลจากแฟ้มข้อมูล HTML พบว่า MUSEARCH สามารถค้นหาข้อมูลให้ได้ข้อมูลตรงกับความต้องการ และค้นหาข้อมูลที่อยู่ในรูปแบบ Mulitmedia ได้ ด้วยความถูกต้องที่น่าพอใจ

Copyright by Mahidol University

Thesis Title	Object-Oriented Modeling Approach To World Wide Web Searching
Name	Pagaporn Ekareejit
Degree	Master of Science (Computer Science)
Thesis Supervisory Committee	Damras Wongsawang, Ph D
	Supachai Tangwongsan, Ph.D.
Date of Graduation	19 December B.E 2539 (1996)

ABSTRACT

The World Wide Web (WWW) searching has, currently, become difficult to retrieve the relevant information because of the overloading information. Furthermore, the WWW structure is a hypertext composed of multimedia data such as text, image, video and audio. So searching with links in HTML file is not appropriate The Internet users need an efficient search engine to alleviate the searching problems.

The MUSEARCH is the prototype model designed for searching the relevant information, including multimedia data, by applying object-oriented modeling approach. Four hierarchical relationships of object-oriented technique : classification, generalization, aggregation and association are applied for modeling of seven object types . TEXT, IMAGE, AUDIO, VIDEO, APPLICATION, JAVAAPPLET and JAVASCRIPT.

The MUSEARCH architecture has four main components: the user interface , the collector engine, the search engine and database The user interface performs the interconnecting between users and collector engine and between users and search engine. The collector engine is the author's information registration. It will generate the object types and the essential details for searching from HTML files. The search engine is used for querying with keywords, subjects, URL and specific object types It returns the correspondent HTML page satisfying the user requirement. The database keeps the data extracted from the collector engine. To improve searching speed, the MUSEARCH uses the indexing of the words called "Inverted Object Vectors".

The experiment with simulating HTML files has been conducted The results show that the MUSEARCH can find multimedia information with also its relevance with satisfactory correctness.

Copyright by Mahidol University