การรวมข้อมูลสารสนเทศจากแหล่งสารสนเทศวิวิธพันธุ์โดยวิธีออนโทโลยี

นางงามนิจ อาจอินทร์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2546

ISBN 974-17-3952-4

ONTOLOGY-BASED APPROACH FOR GATHERING THE HETEROGENEOUS INFORMATION SOURCES

Mrs. Ngamnij Arch-int

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Computer Science
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic Year 2003

ISBN 974-17-3952-4

Thesis Title	Ontology-based Approach for Gathering the Heterogeneous Information Sources
By	Mrs. Ngamnij Arch-int
Field of Study	Computer Science
Thesis Advisor	Assistant Professor Peraphon Sophatsathit, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctor's Degree

..... Dean of Faculty of Science

(Professor Piamsak Menasveta, Ph.D.)

THESIS COMMITTEE

..... Chairman
(Professor Chidchanok Lursinsap, Ph.D.)

..... Thesis Advisor
(Assistant Professor Peraphon Sophatsathit, Ph.D.)

..... Member
(Assistant Professor Pattarasinee Bhattarakosol, Ph.D.)

..... Member
(Assistant Professor Somjai Boonsiri, Ph.D.)

..... Member
(Surapant Meknavin, Ph.D.)

งามนิจ อาจอินทร์ : การรวมข้อมูลสารสนเทศจากแหล่งสารสนเทศวิวิธพันธุ์โดยวิธีออนโทโลยี (Ontology-based Approach for Gathering the Heterogeneous Information Sources)
อ. ที่ปรึกษา: ผศ. ดร. พีระพนธ์ โสพัศสถิตย์, จำนวนหน้า 144 **หน้า**. ISBN 974-17-3952-4

การเข้าถึงและการรวบรวมข้อมูลสารสนเทศจากแหล่งสารสนเทศวิวิธพันธุ์ เป็นปัญหาหลักของระบบการประมวลผลแบบกระจาย ซึ่งสาระสำคัญของปัญหาอันหนึ่ง คือความหลากหลายทางด้านความหมายของข้อมูล หลายระบบถูกนำเสนอเพื่อแก้ปัญหาที่เกิดขึ้น ตั้งแต่ระบบตัวกลาง (Mediator-based Systems) ไปจนถึงระบบวงจรการพรรณนา (Description logic-based systems) อย่างไรก็ตาม วิธีการแก้ปัญหานี้ในปัจจุบัน ยังมีข้อจำกัดทางด้านความยืดหยุ่น ความสามารถในการขยายระบบ การปฏิบัติงานร่วมกัน และความแข็งแกร่งของระบบ

วิทยานิพนธ์นี้ได้นำเสนอ ระบบการรวบรวมข้อมูลสารสนเทศแบบสื่อความหมาย (Semantic Information Gathering Approach) ขึ้น สำหรับการเข้าถึงและการรวบรวมข้อมูลสารสนเทศ จากแหล่งสารสนเทศวิวิธพันธุ์บนเครือข่ายอินเตอร์เน็ต สถาปัตยกรรมของระบบออกแบบตามหลักการของโครงสร้างที่แบ่งระดับชั้นการทำงานเป็นอิสระจากกัน (Layered-architecture) โดยการผนวกซอฟต์แวร์ตัวแทนที่เคลื่อนที่ได้ (Mobile agent) ทำหน้าที่เป็นตัวเชื่อมระหว่าง ลูกข่ายและแม่ข่าย (Client/Server) เพื่อลดภาระงานที่หนักอันเนื่องมาจากสถาปัตยกรรม ลูกข่าย/แม่ข่าย วิทยานิพนธ์นี้ยังนำเสนอ พจนานุกรมข้อมูล (Metadata Dictionary) ซึ่งเป็นองค์ประกอบที่สำคัญของสถาปัตยกรรมข้างต้น เพื่อแก้ปัญหาความหลากหลายทางด้านความหมายของข้อมูล องค์ประกอบของพจนานุกรมข้อมูลนี้ ได้จากออนโทโลยีเฉพาะกิจ (Domain Ontology) ซึ่งนิยามจากหลักการเชิงวัตถุและทฤษฎีเซต เพื่อให้ระบบสามารถปฏิบัติงานร่วมกันได้ดี อันเป็นการประยุกต์ที่เหมาะสมกับการทำงานบนอินเตอร์เน็ต ในส่วนของระบบใช้ภาษา XML เป็นภาษาสำหรับการแสดงพจนานุกรมข้อมูล วิทยานิพนธ์นี้ยังครอบคลุมรายละเอียดของการทำงานทั้งหมด สถาปัตยกรรมของซอฟต์แวร์ตัวแทน กระบวนการสร้างพจนานุกรมข้อมูลบนพื้นฐานของออนโทโลยี องค์ประกอบและการแสดงองค์ประกอบของพจนานุกรม รวมถึงกระบวนการค้นหาข้อมูลจากแหล่งสารสนเทศวิวิธพันธุ์ด้วยพจนานุกรมข้อมูล

ภาควิชา คณิตศาสตร์ **ลายมือชื่อนิสิต** .....
สาขาวิชา วิทยาการคอมพิวเตอร์ **ลายมือชื่ออาจารย์ที่ปรึกษา** .....
ปีการศึกษา 2546

4373810423: MAJOR COMPUTER SCIENCE
KEYWORDS: HETEROGENEOUS INFORMATION SOURCES/ DOMAIN ONTOLOGY/ XML-BASED METADATA DICTIONARY/ QUERY PROCESSING
NGAMNIJ ARCH-INT: ONTOLOGY-BASED APPROACH FOR GATHERING THE HETEROGENEOUS INFORMATION SOURCES. THESIS ADVISOR: ASST. PROF. PERAPHON SOPHATSATHIT, Ph.D., 144 pp. ISBN 974-17-3952-4.

The problem of accessing and integrating heterogeneous information sources has become the center stage in the distributed processing environment. One of the important issues stemming from accessing these heterogeneous sources is semantic heterogeneity. A number of systems have been proposed to address this issue, ranging from mediator-based systems to description logic-based systems. However, the current methodologies for accessing heterogeneous sources offer limited flexibility, scalability, interoperability, and robustness.

This dissertation proposes the Semantic Information Gathering Approach, hereafter *SIGA*, a system for accessing and integrating heterogeneous sources on the WWW. The reference architecture is based on layered-architecture incorporating mobile agents as the client/server connectivity to eliminate traditional client/server overhead. A metadata dictionary which is an essential component of the reference architecture is also proposed for solving semantic heterogeneity. The metadata dictionary is derived from domain ontology where the constituent components are defined in terms of object-oriented principle and set theory. To enable system-wide interoperability suitable for a Web-based environment, the XML technology is selected as the language for expressing the metadata dictionary contents. Consequently, this dissertation also covers the overall system and agent architecture, the modeling process of the ontology-based metadata dictionary, the components and representation of the metadata dictionary, and the querying process in accessing and integrating the heterogeneous sources through the metadata dictionary.

Department of Mathematics Student's signature Ngamnij Arch-int.
Field of study Computer Science Advisor's signature
Academic year 2003

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Peraphon Sophatsathit, for his invaluable guidance, constructive comments, and continuous encouragement throughout the period of this study. Gratitude is also extended to Prof. Chidchanok Lursinsap, Dr. Pattarasinee Bhattarakosol, Dr. Somjai Boonsiri, and Dr. Surapant Meknavin, my examination committee members, for their reading, comments and suggestions in making this dissertation a reality.

I am very grateful to Dr. Yuefeng Li, and Dr. Paul Roe, who provided me the opportunity as a visiting scholar at the School of Software Engineering and Data Communications, Queensland University of Technology, Brisbane, Australia, as well as their suggestions and assistance during my visiting for 8 months.

I also would like to express my gratitude to the Ministry of University Affairs for all financial support throughout my study.

A special thank is given to the members of AVIC and staff for their help and friendship.

Finally, my deepest gratitude goes to my beloved father, mother, husband, daughter and son, for their invaluable love, support, and encouragement throughout my whole life.

TABLE OF CONTENTS

page

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

LIST OF FIGURES

page

LIST OF FIGURES (Cont'd)

LIST OF TABLES

page