

การค้นคืนข้อมูลผ่านทางระบบฟรอนต์เอนด์โดยใช้ประโยคภาษาไทย

นายเกรียงศักดิ์ อัครรัตน์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาศึกษาศาสตร์ มหาวิทยาลัย

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2539

ISBN 974-634-499-1

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I17423946

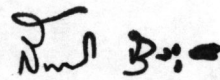
DATA RETRIEVAL VIA THE FRONT-END USING THAI SENTENCES

MR. KREANGSAK ASAVARATANA

A Thesis submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
Department of Computer Engineering
Graduate School
Chulalongkorn University
Academic Year 1996
ISBN 974-634-499-1

หัวข้อวิทยานิพนธ์ การค้นคืนข้อมูลผ่านทางระบบฟรอนต์เอนด์โดยใช้ประโยคภาษาไทย
โดย นาย เกรียงศักดิ์ อัครรัตน์
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา อาจารย์ จารุมาตร ปิ่นทอง

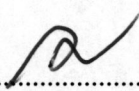
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต



..... คณบดีบัณฑิตวิทยาลัย

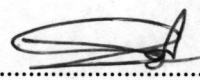
(รองศาสตราจารย์ ดร. สันติ ฤงสุวรรณ)

คณะกรรมการสอบวิทยานิพนธ์



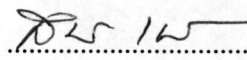
..... ประธานกรรมการ

(รองศาสตราจารย์ เตือน สินธุ์พันธ์ประทุม)



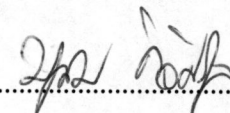
..... อาจารย์ที่ปรึกษา

(อาจารย์ จารุมาตร ปิ่นทอง)



..... อาจารย์ที่ปรึกษาร่วม

(ผู้ช่วยศาสตราจารย์ ดร. กิ่งกาญจน์ เทพกาญจนา)



..... กรรมการ

(อาจารย์ ดร. บุญเสริม กิจศิริกุล)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

เกรียงศักดิ์ อัครรัตน์ : การค้นคืนข้อมูลผ่านทางระบบฟรอนต์เอนด์โดยใช้ประโยคภาษาไทย
(DATA RETRIEVAL VIA THE FRONT-END USING THAI SENTENCES) อ. ที่ปรึกษา
: อ. จารุมাত্র ปันทอง, 73 หน้า. ISBN 974-634-499-1

วิทยานิพนธ์นี้มีจุดประสงค์เพื่อออกแบบและสร้างระบบฟรอนต์เอนด์เพื่อค้นคืนข้อมูลใน
ฐานข้อมูลแบบสัมพันธ์โดยใช้ประโยคภาษาไทย ประโยคภาษาไทยจะถูกแปลงเป็นคำสั่งเอสคิวแอลที่
สามารถนำไปใช้ในการค้นคืนข้อมูลจากฐานข้อมูลได้ ระบบฟรอนต์เอนด์ที่ออกแบบประกอบด้วย 3 ส่วน
คือ (1) ส่วนวิเคราะห์ประโยคประโยคภาษาไทย ใช้เทคนิคการวิเคราะห์กระจายแบบแอลอาร์ในการ
กระจายไวยากรณ์ไม่พึงบริบท และใช้ทฤษฎีเอชพีเอสจีในการวิเคราะห์ประโยคโดยนำเสนอโครงสร้าง
คุณสมบัติในเอชพีเอสจีด้วยกราฟไรวังที่มีทิศทาง (2) ส่วนแปลความหมาย ใช้ฐานความรู้ที่ถูกจัดให้อยู่
ในกรอบของโครงสร้างคุณสมบัติในการช่วยแปลความหมาย โดยใช้ตัวดำเนินการยูนิฟิเคชันเพื่อรวบรวม
ความหมายของประโยค (3) ส่วนสร้างคำสั่งสอบถาม รวบรวมข้อมูลจากโครงสร้างความหมาย และนำ
ข้อมูลเหล่านั้นมาจัดสร้างเป็นภาษาเอสคิวแอล ระบบโปรแกรมนี้เขียนขึ้นมาโดยใช้ภาษา C และ C++ ใช้
กับเครื่องไมโครคอมพิวเตอร์บนระบบปฏิบัติการวินโดวส์ ผลการทดลองโดยใช้ฐานข้อมูลห้องสมุดเป็น
ตัวอย่างสามารถใช้ประโยคภาษาไทยในการค้นคืนข้อมูลจากฐานข้อมูลได้ตามวัตถุประสงค์

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2539

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

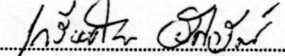
C518769 : MAJOR COMPUTER ENGINEERING
KEY WORD: : FRONT-END / DATA RETRIEVAL / NATURAL LANGUAGE PROCESSING
KREANGSAK ASAVARATANA : DATA RETRIEVAL VIA THE FRONT-END USING
THAI SENTENCES. THESIS ADVISOR : CHARUMATR PINTHONG. 73 pp.
ISBN 974-634-499-1

The purpose of this thesis is to illustrate and build a front-end system for relational database retrieval using Thai sentences. Thai sentence, process by the system, will be translated to SQL command for data retrieval. The front-end system is composed of 3 parts. (1) Parser, using LR-parsing technique to parse CFG, HPSG for sentence parsing, and representing feature structure as directed acyclic graph. (2) Semantic interpreter, using knowledge-base in frame work of feature structure to interpret and unification operator to constitute structure. (3) Query generator, integrating of information of the structure and then generating a SQL command. The system, implemented by C and C++, runs on a micro-computer under Microsoft Windows Operating System. The result, demonstrated with a sample library database, is able to use Thai sentences for database retrieval.

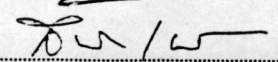
ภาควิชา..... วิศวกรรมคอมพิวเตอร์

สาขาวิชา..... วิศวกรรมคอมพิวเตอร์

ปีการศึกษา..... 2539

ลายมือชื่อนิสิต..... 

ลายมือชื่ออาจารย์ที่ปรึกษา..... 

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม..... 



กิตติกรรมประกาศ

ในการจัดทำวิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ อาจารย์ จารุมาตร ปิ่นทอง และ ผู้ช่วยศาสตราจารย์ ดร. กิ่งกาญจน์ เทพกาญจนา เป็นอย่างสูงที่ได้ให้โอกาสแก่ผู้วิจัยจัดทำวิทยานิพนธ์หัวข้อนี้ ตลอดจนให้ความกรุณาและติดตามผลการวิจัย รวมทั้งให้คำปรึกษาและคำแนะนำต่างๆเกี่ยวกับการทำวิทยานิพนธ์มาโดยตลอด

ขอขอบพระคุณครอบครัวของข้าพเจ้าที่ให้การสนับสนุน และเป็นกำลังใจในการศึกษาครั้งนี้ และขอขอบพระคุณ อาจารย์และเจ้าหน้าที่แห่งภาควิชาวิศวกรรมคอมพิวเตอร์ที่อำนวยความสะดวกในการทำวิทยานิพนธ์

ขอขอบคุณ เพื่อนๆทุกคนที่ให้ความช่วยเหลือ คำแนะนำ และความร่วมมือเป็นอย่างดีในการทำวิทยานิพนธ์นี้

สารบัญ

หน้า

บทคัดย่อไทย.....	ง
บทคัดย่ออังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์.....	3
ขอบเขตของการวิจัย.....	3
ขั้นตอนของการวิจัย.....	4
ประโยชน์ที่ได้จากการวิจัย.....	5
บทที่ 2 การแปลงประโยคภาษาไทยไปเป็นคำสั่งสอบถาม.....	6
การวิเคราะห์ประโยคภาษาไทย.....	7
การแปลความหมาย.....	9
การสร้างคำสั่งสอบถาม.....	9
บทที่ 3 การแบ่งคำ.....	11

พยางค์ และ คำ	12
ยุทธวิธีพื้นฐานของการแบ่งคำ.....	12
ทฤษฎีการแบ่งคำด้วยพจนานุกรม.....	14
อัลกอริทึมแบ่งคำด้วยพจนานุกรม	21
การแบ่งคำด้วยพจนานุกรมกับการวิเคราะห์กระจายประโยค	21
บทที่ 4 การวิเคราะห์กระจายประโยค.....	23
การทำให้เกิดผลของโครงสร้างคุณสมบัติ	23
การนำเสนอคลังคำเอชพีเอสจี	25
อัลกอริทึมวิเคราะห์กระจาย	26
การประกอบเป็นโครงสร้างองค์ประกอบ.....	29
ข้อผิดพลาดในการวิเคราะห์กระจาย	30
บทที่ 5 การแปลความหมายและการสร้างคำสั่งสอบถาม.....	31
การแปลความหมาย.....	31
อัลกอริทึมแปลความหมาย.....	33
การสร้างคำสั่งสอบถาม	34
อัลกอริทึมสร้างภาษาเอสคิวแอล	35
บทที่ 6 การทำงานของโปรแกรม.....	37
ส่วนประกอบของโปรแกรม.....	37
ขั้นตอนการทำงาน.....	39
บทที่ 7 สรุปการวิจัยและข้อเสนอแนะ.....	44
ข้อเสนอแนะ	45
รายการอ้างอิง	46

ภาคผนวก..... 48

ภาคผนวก ก ตัวอย่างประโยคและผลของการประมวลผล 49

ประวัติผู้เขียน..... 62

สารบัญตาราง

หน้า

ตารางที่ 4.1 ตารางไวยากรณ์ไม่พึงบริบทสำหรับตัวกระจายไวยากรณ์แบบแอลอาร์.....	28
---	----

สารบัญญรูปภาพ

	หน้า
รูปที่ 2.1 ส่วนประกอบของตัวแปลงประโยคเป็นคำสั่งสอบถาม.....	6
รูปที่ 3.1 DFA ของพจนานุกรม D ในรูปของ trie.....	17
รูปที่ 3.2 NFA ของภาษา S	18
รูปที่ 4.1 โครงสร้างคุณสมบัติที่ถูกแสดงแทนด้วยโครงสร้างกราฟ.....	24
รูปที่ 4.2 ความสัมพันธ์ของลำดับขั้นการถ่ายทอดคุณสมบัติของ Value และ ValuePtr	25
รูปที่ 4.4 แสดงส่วนประกอบในการวิเคราะห์ประโยค.....	27
รูปที่ 5.2 รหัสเทียมของอัลกอริทึมแปลความหมาย.....	34
รูปที่ 5.3 รหัสเทียมของโมดูลค้นหาแบบโลคอล.....	35
รูปที่ 5.4 รหัสเทียมของโมดูลค้นหาแบบโกลบอล	36
รูปที่ 6.1 แสดงหน้าที่ของแต่ละแฟ้มข้อมูลของโปรแกรม.....	38
รูปที่ 6.2 โครงสร้างของระบบหลังจากการดัดแปลงแล้ว	39
รูปที่ 6.3 แสดงความสัมพันธ์ของฐานข้อมูลจำลองของห้องสมุด.....	40
รูปที่ 6.4 แสดงหน้าจอของโปรแกรมเมื่อเริ่มต้นทำงาน.....	41
รูปที่ 6.5 ผลลัพธ์ของการแปลงเป็นประโยค “มีหนังสือกี่เล่ม” เป็นเอสคิวแอล	41
รูปที่ 6.6 แสดงโครงสร้างคุณสมบัติของประโยค “มีหนังสือกี่เล่ม”	42
รูปที่ 6.7 แสดงโครงสร้างคุณสมบัติที่เป็นรหัสชั่วคราวของประโยค “มีหนังสือกี่เล่ม”	42
รูปที่ 6.8 ผลลัพธ์ของการแปลงประโยค “แสดงจำนวนของหนังสือ” เป็นเอสคิวแอล	43