

การออกแบบและพัฒนาขั้นตอนวิธีการวิเคราะห์บันทึกการเข้าใช้โดยใช้เทคนิคการทำเหมืองข้อมูล



นาย วิญญู ละอองสุวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2544

ISBN 974-03-0774-4

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DESIGN AND DEVELOPMENT OF A METHOD FOR WEB ACCESS LOG ANALYSIS
USING DATAMINING TECHNIQUES

Mr. Vinyoo La-ongsuwan

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2001

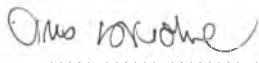
ISBN 974-03-0774-4

หัวข้อวิทยานิพนธ์ การออกแบบและพัฒนาขั้นตอนวิธีการวิเคราะห์บันทึกการเข้าใช้โดยใช้
เทคนิคการทำเหมืองข้อมูล
โดย นาย วิชาญ ละอองสุวรรณ
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล


คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโท

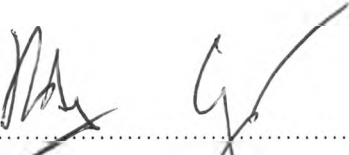

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์


.....ประธานกรรมการ
(อาจารย์ ดร.ยรรยง เต็งอำนวย)


.....อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)


.....กรรมการ
(อาจารย์ ชัยศิริ ปันนิตานนท์)


.....กรรมการ
(อาจารย์ ดร. ณัฐวดี นุไฟโรจน์)

วิญญู ละอองสุวรรณ : การออกแบบและพัฒนาขั้นตอนวิธีการวิเคราะห์บันทึกการเข้าใช้
โดยใช้เทคนิคการทำเหมืองข้อมูล. (DESIGN AND DEVELOPMENT OF A METHOD FOR
WEB ACCESS LOG ANALYSIS USING DATAMINING TECHNIQUES)

อ.ที่ปรึกษา : ผศ.ดร.บุญเสริม กิจศิริกุล, 84 หน้า. ISBN 974-03-0774-4.

งานวิจัยนี้นำเสนอขั้นตอนวิธีการค้นพบรูปแบบการเยี่ยมชมไซต์ของผู้ชมซึ่งเป็นส่วนหนึ่ง
ของการทำเหมืองข้อมูลการใช้งานเว็บไซต์ ซึ่งเป็นการประยุกต์ใช้งานของการทำเหมืองข้อมูลและ
เทคนิคการค้นพบความรู้บนเว็ลด์ไวด์เว็บทรานแซคชัน เจนำเสนอแบบจำลองสำหรับงานทางการ
ทำเหมืองข้อมูลการใช้งานเว็บไซต์ ได้แก่การค้นพบกฎความสัมพันธ์และรูปแบบลำดับการเยี่ยมชม
ชมเว็บไซต์ นอกจากนี้เรายังนำเสนอระบบการทำเหมืองข้อมูลการใช้งานเว็บไซต์ซึ่งสร้างอยู่บนชั้น
ตอนวิธีที่นำเสนอและอธิบายถึงผลการทดลองบนข้อมูลจริงอีกด้วย

ภาควิชา.....วิศวกรรมคอมพิวเตอร์.....

สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์.....

ปีการศึกษา.....2544.....

ลายมือชื่อนิสิต.....วิญญู ละอองสุวรรณ.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....ผศ.ดร.บุญเสริม กิจศิริกุล.....

4270534621 : MAJOR COMPUTER SCIENCE

KEY WORD: DATA MINING / KNOWLEDGE DISCOVERY / WORLD WIDE
WEB / ASSOCIATION RULES / SEQUENTIAL PATTERNS / WEB MINING

VINYOO LA_ONGSUWAN : DESIGN AND DEVELOPMENT OF A METHOD FOR WEB
ACCESS LOG ANALYSIS USING DATAMINING TECHNIQUES.

THESIS ADVISOR : Assoc. Prof. BOONSERM KIJSIRIKUL, Ph.D.,

84 pp. ISBN 974-03-0774-4.


This thesis proposes a method for web usage mining that is the application of data mining and knowledge discovery techniques to mine data in world wide web transactions. We present models for various web usage mining tasks such as the discovery of association rules and sequential patterns from web data. We also present a web usage mining system which has been implemented based upon the proposed method and discuss our experimental results on real web data.

Department.....Computer Engineering.....

Fields of study...Computer Science.....

Academic year...2001.....

Student's signature..........

Advisor's signature..........

Co-Advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของอาจารย์ที่ปรึกษา
วิทยานิพนธ์ ดร.บุญเสริม กิจศิริกุล ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ในการวิจัยมาด้วย
ดีตลอด รวมทั้งตรวจแก้วิทยานิพนธ์ฉบับนี้อย่างละเอียด ผู้เขียนขอขอบพระคุณท่านอาจารย์ภาค
วิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ประสิทธิประสาทวิชา

ขอขอบคุณสมาชิกห้องปฏิบัติการอัจฉริยภาพเครื่องกลและการค้นพบความรู้ (MIND
LAB) ที่ให้ความรู้และข้อคิดเห็นที่ดี ๆ และเป็นประโยชน์ต่อการวิจัย รวมทั้งเหล่าเพื่อน พี่ น้อง ร่วม
รุ่นที่ช่วยเหลือทั้งในด้านวิชาการ สันทนาการ กิจกรรมพิเศษในระหว่างที่ศึกษาและทำงานวิจัยจน
สำเร็จลุล่วงมาได้ด้วยดี

ท้ายนี้ ผู้วิจัยใคร่ขอขอบพระคุณบิดา มารดา ที่คอยสนับสนุน ฝ่ากระดุนถามความคืบ
หน้า และให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฌ
สารบัญภาพ	ญ
บทที่	
1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	2
1.4 ขั้นตอนและวิธีการดำเนินการวิจัย	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
1.6 โครงสร้างของวิทยานิพนธ์	3
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 โครงสร้างข้อมูลของล็อกไฟล์ของเว็บเซิร์ฟเวอร์	4
2.2 แพทริกเกอร์	6
2.3 การเรียนรู้กฎความสัมพันธ์	6
2.4 การเรียนรู้รูปแบบลำดับ	9
2.5 งานวิจัย WEBMINER	11
2.6 โปรแกรมวิเคราะห์บันทึกการเข้าใช้เว็บไซต์	12
3 ขั้นตอนวิธีการประมวลผลข้อมูลเบื้องต้น	14
3.1 ขั้นตอนการเก็บรวบรวมข้อมูล	14
3.2 ขั้นตอนการกำหนดรหัสประจำตัวผู้ชม	14
3.3 ขั้นตอนการกำหนดทรานเซคชัน	16
3.4 ขั้นตอนการวิเคราะห์ข้อมูลทางสถิติ	17
3.5 สรุป	18

สารบัญ (ต่อ)

4	ขั้นตอนวิธีการเรียนรู้กฎความสัมพันธ์และรูปแบบลำดับ.....	19
4.1	ขั้นตอนวิธีการเรียนรู้กฎความสัมพันธ์.....	19
4.2	ขั้นตอนวิธีการเรียนรู้รูปแบบลำดับ.....	25
4.3	สรุป.....	31
5	การทดลองและผลการทดลอง.....	33
5.1	วิธีการทดลอง.....	33
5.2	ผลการทดลอง.....	34
5.3	วิธีการนำกฎความสัมพันธ์และรูปแบบลำดับไปใช้ปรับโครงสร้างลิงก์ของเว็บไซต์.....	43
5.4	สรุป.....	46
6	สรุปผลการวิจัยและข้อเสนอแนะ.....	47
6.1	สรุปผลการวิจัย.....	47
6.2	ข้อดีและข้อเสียของขั้นตอนวิธี.....	47
6.3	ข้อเสนอแนะ.....	48
	รายการอ้างอิง.....	50
	ภาคผนวก.....	51
	ภาคผนวก ก.....	52
	ภาคผนวก ข.....	54
	ประวัติผู้วิจัย.....	75

สารบัญญัตราสาร

	หน้า
ตารางที่ 1 ฟิลด์ของล็อกไฟล์แบบคอมไบนี.....	4
ตารางที่ 2 ความหมายของพารามิเตอร์.....	34
ตารางที่ 3 ชื่อชุดข้อมูลและพารามิเตอร์ที่ใช้ในการสร้างข้อมูลทดสอบ	35
ตารางที่ 4 ความสัมพันธ์ระหว่างค่าซีพอร์ตขึ้นต่ำกับจำนวนกฎความสัมพันธ์	37
ตารางที่ 5 ความสัมพันธ์ระหว่างค่าความเชื่อมั่นขึ้นต่ำกับจำนวนกฎความสัมพันธ์	38
ตารางที่ 6 ความหมายของพารามิเตอร์.....	38
ตารางที่ 7 ชื่อชุดข้อมูลและพารามิเตอร์ที่ใช้ในการสร้างข้อมูลทดสอบ	39
ตารางที่ 8 ความสัมพันธ์ระหว่างค่าซีพอร์ตขึ้นต่ำกับจำนวนรูปแบบลำดับ	41
ตารางที่ 9 ความสัมพันธ์ระหว่างค่าความเชื่อมั่นขึ้นต่ำกับจำนวนรูปแบบลำดับ	42
ตารางที่ 10 ตัวอย่างกฎความสัมพันธ์ที่ค้นพบ.....	43
ตารางที่ 11 ตัวอย่างรูปแบบลำดับที่ค้นพบ	44

สารบัญภาพ

	หน้า
รูปที่ 1 ภาพของล็อกไฟล์แบบคอมมอน.....	5
รูปที่ 2 ภาพของล็อกไฟล์แบบคอมไบนี.....	5
รูปที่ 3 สถาปัตยกรรมของ WEBMINER.....	12
รูปที่ 4 อัลกอริทึม APRIORI.....	21
รูปที่ 5 ฟังก์ชัน apriori-gen (join step)	21
รูปที่ 6 ฟังก์ชัน apriori-gen (prune step).....	22
รูปที่ 7 ตัวอย่างการทำงานของอัลกอริทึม APRIORI	22
รูปที่ 8 ฟังก์ชัน genrules	24
รูปที่ 9 อัลกอริทึม Apriori All	27
รูปที่ 10 ฟังก์ชัน aprioriall-gen (join step).....	28
รูปที่ 11 ฟังก์ชัน aprioriall-gen (prune step).....	28
รูปที่ 12 ตัวอย่างการทำงานของอัลกอริทึม Apriori All.....	29
รูปที่ 13 ฟังก์ชัน genpatterns.....	31
รูปที่ 14 โครงสร้างของโปรแกรม.....	33
รูปที่ 15 เวลาประมวลผลของอัลกอริทึม APRIORI โดยใช้ข้อมูลทดสอบต่างๆ.....	36
รูปที่ 16 เวลาประมวลผลของอัลกอริทึม APRIORI โดยใช้ข้อมูลจากตารางที่ 4	37
รูปที่ 17 เวลาประมวลผลของอัลกอริทึม Apriori All โดยใช้ข้อมูลทดสอบต่างๆ.....	40
รูปที่ 18 เวลาประมวลผลตามจำนวนผู้ชม.....	41
รูปที่ 19 เวลาประมวลผลของอัลกอริทึม Apriori All โดยใช้ข้อมูลจากตารางที่ 8	42