



WEB MINING

SUTTHISAK INTHA WADEE

อภิรักษ์ ินทาวดี

จาก

บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE (COMPUTER SCIENCE)
FACULTY OF GRADUATED STUDIES
MAHIDOL UNIVERSITY**

2001

ISBN 974-04-0570-3

COPY RIGHT OF MAHIDOL UNIVERSITY

Copyright by Mahidol University

TH
S967 W
2001

3937012 SCCS/M

: MAJOR : COMPUTER SCIENCE ; MSc.
(COMPUTER SCIENCE)

KEY WORDS

: DATA MINING / WEB MINING / WEB DATA /
USAGE PATTERN / FUZZY SYSTEMSUTTHISAK INTHAWADEE : WEB MINING. THESIS ADVISORS:
DAMRAS WONGSAWANG, Ph.D., THANWADEE SUNETNANTA, Ph.D. 130p.
ISBN 974-04-0570-3

Web mining is an application of data mining techniques to retrieve knowledge from the huge Web data. The term Web mining has been used in two distinct ways: Web Content Mining and Web Usage Mining. In this thesis, Web Usage Mining is the main area of study. The objective of Web Mining is to discover and extract usage patterns from the Web data. The objective of this study is to develop a new model of Web usage mining which can retrieve valuable information of usage patterns close to the way human beings think. It should also work well in real applications. This thesis proposed a new qualitative model to extract usage patterns from Web data based on users' interest. The new model increases access to areas of subjective interest.

The fuzzy rules based system and association rule mining are the main techniques that were applied to develop the new model. The fuzzy rules based system was applied to find a list of important Web pages for each user. The usage time and usage frequencies are used as the main variables to generate a rules based system. The rules based system helps the user assess the importance of the page. Then a list of Web pages of relative importance is generated. Using the list, the association rule mining is used to find the sets of associated Web pages. The results of both techniques are used to synthesize the Web index on the first Web page. The Web index will help the user to access the pages of interest and associated pages within one step. In this study, the new model is tested with test data gathered from the Web site. <http://www.BuyOnlineNow.Com>. The results of the tests proved that the new model can reduce the time to get access to data of interest.

In real applications, the new model can be applied to the B2C web site in order to improve services. It can help in marketing analysis and Web personalization. However, this new model has limitations. It does not work well with dynamic Web sites. Any future study should improve the model so that it can work with dynamic Web sites.

3937012 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์ ; วท.ม. (วิทยาการ
คอมพิวเตอร์)

ผู้ประพันธ์ : อินทวดี : การทำเหมืองข้อมูลบนเว็บ (WEB MINING) คณะกรรมการ
ควบคุมวิทยานิพนธ์ : คำรัส วงศ์สว่าง, Ph.D., ธันวดี สุเนตนันท์, Ph.D., 130 หน้า. ISBN
974-04-0570-3

การทำเหมืองข้อมูลบนเว็บ (Web Mining) เป็นการนำเทคนิคการทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้ในการสืบค้นข้อมูลที่เป็นองค์ความรู้ จากข้อมูลเว็บ ซึ่งมีจำนวนมากมาย
มหาศาล การทำเหมืองข้อมูลบนเว็บมีการศึกษาใน 2 แนวทางคือ การทำเหมืองเนื้อหาข้อมูลบนเว็บ
และการทำเหมืองข้อมูลการใช้งานของผู้ใช้บนเว็บ โดยในการจัดทำวิทยานิพนธ์นี้จะเน้นการศึกษา
การทำเหมืองข้อมูลการใช้งานของผู้ใช้บนเว็บเป็นหลัก โดยวัตถุประสงค์การศึกษาเพื่อค้นหา
วิธีการใหม่ในการดึงรูปแบบการใช้งานของผู้ใช้เว็บ โดยมุ่งเน้นการพัฒนาวิธีการใหม่ที่สามารถสืบ
ค้นข้อมูลรูปแบบการใช้งานเว็บที่มีประโยชน์ และใกล้เคียงกับความถึกมนุษย์ รวมทั้งสามารถนำมา
ใช้งานจริงได้ ซึ่งวิทยานิพนธ์ฉบับนี้จะทำการนำเสนอวิธีการใหม่ในการการทำเหมืองข้อมูลการใช้งาน
ของผู้ใช้บนเว็บเชิงคุณภาพคดขยอาศัยความสนใจผู้ใช้ วิธีการใหม่จะปรับปรุงการวัดความสนใจเชิง
วัตถุประสงค์

ในการพัฒนาวิธีการใหม่ได้ประยุกต์เทคโนโลยีปัญญาประดิษฐ์ที่เรียกว่า Fuzzy Rules
Based System และเทคนิคการทำเหมืองข้อมูล Association Rule Mining ซึ่ง Fuzzy Rules Based
System ถูกประยุกต์ใช้ในการค้นหาความรายการของหน้าเว็บที่สนใจของผู้ใช้ โดยใช้เวลาในการ
ใช้งานประกอบกับความถี่ในการใช้งานในการสร้าง Rules Based System ส่วน Association Rule
Mining ถูกใช้ในการหากลุ่มของเว็บที่มีความเชื่อมโยงกัน ซึ่งเหล่านี้ใช้ในการจัดสร้าง Web Index
ในหน้าเว็บแรกเพื่อให้ผู้ใช้สามารถเข้าถึงเว็บที่มีความสนใจหรือเกี่ยวข้องกันในครั้งเดียว โดยชุดข้อ
มูลที่ทดสอบเก็บจากการใช้งานของผู้ใช้บนเว็บ <http://www.BuyOnlineNow.Com>. ซึ่งจากผลการ
ทดสอบพบว่าวิธีการที่นำเสนอสามารถลดเวลาในการเข้าถึงเว็บที่สนใจได้เป็นอย่างดี

สำหรับในการใช้งานจริงวิธีที่นำเสนอจะเข้าไปช่วยปรับปรุงระบบเว็บเพื่อการค้าในการ
วิเคราะห์ข้อมูลทางการตลาด และการจัดเตรียมข้อมูลบนเว็บที่นำเสนอให้ตรงกับความต้องการผู้ใช้
อย่างไรก็ตามวิธีที่นำเสนอยังมีข้อจำกัดคือไม่สามารถทำงานได้กับเว็บแบบ Dynamic ซึ่ง
ต้องทำการศึกษาหาวิธีแก้ไขข้อจำกัดดังกล่าวในการศึกษาขั้นต่อไป