# Web Mining: Applications and Techniques

## Table of Contents

Preface ........................................................................................................................................................... vi

### Section I: Introduction

**Chapter I. Metadata Management: A Requirement for Web Warehousing and Knowledge Management** .......................................................................................................................... 1  
Gilbert W. Laware, Purdue University, USA

**Chapter II. Mining for Web Personalization** ....................................................................................... 27  
Penelope Markellou, University of Patras, Greece and Research Academic Computer Technology Institute, Greece  
Maria Rigou, University of Patras, Greece and Research Academic Computer Technology Institute, Greece  
Spiros Sirmakessis, University of Patras, Greece and Research Academic Computer Technology Institute, Greece

### Section II: Content Mining

**Chapter III. Using Context Information to Build a Topic-Specific Crawling System** .......................... 50  
Fan Wu, National Chung-Cheng University, Taiwan  
Ching-Chi Hsu, Institute for Information Industry, Taiwan

**Chapter IV. Ontology Learning from a Domain Web Corpus** ............................................................... 69  
Roberto Navigli, University of Rome “La Sapienza”, Italy

**Chapter V. MARS: Multiplicative Adaptive Refinement Web Search** ............................................. 99  
Xiannong Meng, Bucknell University, USA  
Zhixiang Chen, The University of Texas - Pan American, USA
Chapter VI. Exploiting Captions for Web Data Mining ................................. 119
   Neil C. Rowe, U.S. Naval Postgraduate School, USA

Chapter VII. Towards a Danger Theory Inspired Artificial Immune System for Web Mining ................................................................. 145
   Andrew Secker, University of Kent, UK
   Alex A. Freitas, University of Kent, UK
   Jon Timmis, University of Kent, UK

Chapter VIII. XML Semantics .................................................................... 169
   Yasser Kotb, Ain Shams University, Egypt
   Katsuhiko Gondow, Tokyo Institute of Technology, Japan
   Takuya Katayama, Japan Advanced Institute of Science and Technology, Japan

Chapter IX. Classification on Top of Data Cube ........................................ 189
   Lixin Fu, University of North Carolina, Greensboro, USA

Section III: Structure Mining

Chapter X. Data Cleansing and Validation for Multiple Site Link Structure Analysis ................................................................. 208
   Mike Thelwall, University of Wolverhampton, UK

Chapter XI. Extracting and Customizing Information Using Multi-Agents ...... 228
   Mohamed Salah Hamdi, UAE University, United Arab Emirates

Chapter XII. Web Graph Clustering for Displays and Navigation of Cyberspace .. 253
   Xiaodi Huang, The University of Southern Queensland, Australia
   Wei Lai, Swinburne University of Technology, Australia

Section IV: Usage Mining

Chapter XIII. Integrating Semantic Knowledge with Web Usage Mining for Personalization ................................................................. 276
   Honghua Dai, DePaul University, USA
   Bamshad Mobasher, DePaul University, USA

Chapter XIV. Web Usage Mining in Search Engines .................................... 307
   Ricardo Baeza-Yates, Universidad de Chile, Chile

Chapter XV. Efficient Web Mining for Traversal Path Patterns .................... 322
   Zhixiang Chen, The University of Texas - Pan American, USA
   Richard H. Fowler, The University of Texas - Pan American, USA
   Ada Wai-Chee Fu, The Chinese University of Hong Kong, Hong Kong
   Chunyue Wang, The University of Texas - Pan American, USA
Chapter XVI. Analysis of Document Viewing Patterns of Web Search Engine Users ... 339
Bernard J. Jansen, The Pennsylvania State University, USA
Amanda Spink, University of Pittsburgh, USA

Chapter XVII. A Java Technology Based Distributed Software Architecture for Web Usage Mining ... 355
Juan M. Hernansáez, University of Murcia, Spain
Juan A. Botía, University of Murcia, Spain
Antonio F.G. Skarmeta, University of Murcia, Spain

Chapter XVIII. Web Usage Mining: Algorithms and Results ... 373
Yew-Kwong Woon, Nanyang Technological University, Singapore
Wee-Keong Ng, Nanyang Technological University, Singapore
Ee-Peng Lim, Nanyang Technological University, Singapore

Section V: Conclusion

Chapter XIX. The Scent of a Newsgroup: Providing Personalized Access to Usenet Sites through Web Mining ... 393
Giuseppe Manco, Italian National Research Council, Italy
Riccardo Ortale, University of Calabria, Italy
Andrea Tagarelli, University of Calabria, Italy

About the Authors ... 415

Index ... 424