XML Data Mining: Models, Methods, and Applications

Andrea Tagarelli
University of Calabria, Italy
Table of Contents

Foreword........................................................................................................................................viii
Preface................................................................................................................................................ix
Acknowledgment................................................................................................................................. xv

Section 1
Models and Measures

Chapter 1
A Study of XML Models for Data Mining: Representations, Methods, and Issues ...................... 1
Sangeetha Kutty, Queensland University of Technology, Australia
Richi Nayak, Queensland University of Technology, Australia
Tien Tran, Queensland University of Technology, Australia

Chapter 2
Modeling, Querying, and Mining Uncertain XML Data ................................................................. 29
Evgeny Kharlamov, Free University of Bozen-Bolzano, Italy & INRIA Saclay, France
Pierre Senellart, Télécom ParisTech, France

Chapter 3
XML Similarity Detection and Measures ....................................................................................... 53
Sanjay Kumar Madria, Missouri University of Science and Technology, USA
Waraporn Viyanon, Missouri University of Science and Technology, USA

Chapter 4
Efficient Identification of Similar XML Fragments Based on Tree Edit Distance ....................... 78
Hongzhi Wang, Harbin Institute of Technology, China
Jianzhong Li, Harbin Institute of Technology, China
Fei Li, Harbin Institute of Technology, China
Section 2
Clustering and Classification

Chapter 5
Approximate Matching Between XML Documents and Schemas with Applications in XML Classification and Clustering .............................................................. 99
Guangming Xing, Western Kentucky University, USA

Chapter 6
The Role of Schema and Document Matchings in XML Source Clustering ....................................................... 125
Pasquale De Meo, University of Messina, Italy
Giacomo Fiumara, University of Messina, Italy
Antonino Nocera, University Mediterranea of Reggio Calabria, Italy
Domenico Ursino, University Mediterranea of Reggio Calabria, Italy

Chapter 7
XML Document Clustering: An Algorithmic Perspective .............................................................. 154
Panagiotis Antonellis, University of Patras, Greece

Chapter 8
Fuzzy Approaches to Clustering XML Structures .................................................................................. 177
Michal Kozielski, Silesian University of Technology, Poland

Chapter 9
XML Tree Classification on Evolving Data Streams .................................................................................. 199
Albert Bifet, University of Waikato, New Zealand
Ricard Gavalda, UPC Barcelona Tech, Spain

Chapter 10
Data Driven Encoding of Structures and Link Predictions in Large XML Document Collections ................................................................................................................. 219
Markus Hagenbuchner, University of Wollongong, Australia
Chung Tsoi, Macau University of Science and Technology, China
Shu Jia Zhang, University of Wollongong, Australia
Milly Kc, University of Wollongong, Australia

Section 3
Association Mining

Chapter 11
Frequent Pattern Discovery and Association Rule Mining of XML Data .................................................. 243
Qin Ding, East Carolina University, USA
Gnanasekaran Sundarraj, Pennsylvania State University, USA
Chapter 12
A Framework for Mining and Querying Summarized XML Data through Tree-Based Association Rules ......................................................... 264
  Mirjana Mazuran, Politecnico di Milano, Italy
  Elisa Quintarelli, Politecnico di Milano, Italy
  Angelo Rauseo, Politecnico di Milano, Italy
  Letizia Tanca, Politecnico di Milano, Italy

Chapter 13
Discovering Higher Level Correlations from XML Data ........................................ 288
  Luca Cagliero, Politecnico di Torino, Italy
  Tania Cerquitelli, Politecnico di Torino, Italy
  Paolo Garza, Politecnico di Milano, Italy

Section 4
Semantics-Aware Mining

Chapter 14
XML Mining for Semantic Web ................................................................. 317
  Rafael Berlanga, Universitat Jaume I, Spain
  Victoria Nebot, Universitat Jaume I, Spain

Chapter 15
A Component-Based Framework for the Integration and Exploration of XML Sources .... 343
  Pasquale De Meo, University of Messina, Italy
  Antonino Nocera, University Mediterranea of Reggio Calabria, Italy
  Domenico Ursino, University Mediterranea of Reggio Calabria, Italy

Chapter 16
Matching XML Documents at Structural and Conceptual Level using Subtree Patterns .................................. 378
  Qi Hua Pan, Curtin University, Australia
  Fedja Hadzic, Curtin University, Australia
  Tharam S. Dillon, Curtin University, Australia

Section 5
Applications

Chapter 17
Geographical Map Annotation with Significant Tags Available from Social Networks ........ 425
  Elena Roglia, University of Turin, Italy
  Rosa Meo, University of Turin, Italy
  Enrico Ponassi, University of Turin, Italy
Chapter 18
Organizing XML Documents on a Peer-to-Peer Network by Collaborative Clustering

Francesco Gullo, University of Calabria, Italy
Giovanni Ponti, ENEA, Italy
Sergio Greco, University of Calabria, Italy

Chapter 19
Incorporating Qualitative Information for Credit Risk Assessment through Frequent Subtree Mining for XML

Novita Ikasari, Curtin University, Australia & University of Indonesia, Indonesia
Fedja Hadzic, Curtin University, Australia
Tharam S. Dillon, Curtin University, Australia

About the Contributors

Index