Contents

Abstract **7**

Foreword **9**

Preface **11**

I The World of the New **15**

1 Web 2.0 Revolution **17**
1.1 What is Web 2.0? 17
1.2 New forms of participation — push or pull? 21
1.3 New forms of expression — blogs 22
1.4 New forms of conversation — Internet forums 23
1.5 New forms of trade — online auctions 25
1.6 New forms of data — mobile objects 28
1.7 Introduction to data mining 29
1.8 Main thesis of the dissertation 32

2 Social-Driven Data **35**
2.1 Introduction 35
2.2 Data from blogs 39
2.3 Data from Internet forums 44
2.4 Data from online auctions 47
2.5 Social implications of the Web 2.0 revolution 52

II Mining of the New **57**

3 Blogosphere **59**
3.1 Introduction 59
3.2 Related Work 61
3.3 Basic Definitions 64
3.4 Trendoo Algorithm 67

3.5 Experiments 74
3.6 Conclusions 79

4 Internet Forums **81**
4.1 Crawling Internet forums 81
4.2 Statistical analysis 83
4.2.1 Topic statistics 83
4.2.2 Post statistics 86
4.2.3 User statistics 87
4.3 Index analysis 91
4.4 Network analysis 99
4.4.1 Model of Internet forum sociogram 99
4.4.2 Topic analysis 102
4.4.3 User analysis 104
4.4.4 Role analysis 106
4.5 Conclusions 107

5 Online Auctions **109**
5.1 Introduction 109
5.2 Related work 113
5.3 Credibility 115
5.3.1 Basic Definitions 116
5.3.2 CredMine Algorithm 117
5.3.3 Experiments 118
5.3.4 Conclusions 121
5.4 Density 123
5.4.1 Basic Definitions 124
5.4.2 Experiments 126
5.4.3 Conclusions 134
5.5 Implicit feedback 135
5.5.1 Existence of Implicit Feedback 135
5.5.2 Simulation 138
5.5.3 Experiments 142
5.5.4 Conclusions 144
5.6 Positive and Negative Reputation 144
5.6.1 Basic Definitions 145
5.6.2 R^+ and R^- Algorithms 146
5.6.3 Experiments 148
5.6.4 Conclusions 153
5.7 Summary of online auction mining 154

III Miscellaneous **155**

6 Moving Objects **157**
6.1 Introduction . . . 158
6.2 Related Work . . . 159
6.3 Basic Definitions . . . 160
6.4 Algorithms . . . 163
6.4.1 AprioriTraj . . . 163
6.4.2 Traj-PrefixSpan . . . 167
6.5 Experiments . . . 169
6.5.1 AprioriTraj . . . 169
6.5.2 Traj-PrefixSpan . . . 171
6.6 Conclusions . . . 176

7 Negative Patterns **179**
7.1 Introduction . . . 180
7.2 Related Work . . . 181
7.3 Basic Definitions . . . 182
7.3.1 Frequent Itemsets and Association Rules . . . 182
7.3.2 Dissociation Itemsets and Dissociation Rules . . . 184
7.4 Algorithms . . . 185
7.5 Experiments . . . 189
7.6 Conclusions . . . 192

8 Summary **195**

IV Appendixes **201**

A Trendoo **203**
A.1 Introduction . . . 203
A.2 Architecture . . . 204
A.3 User Guide . . . 207

B Foruminer **209**
B.1 Introduction . . . 209
B.2 Architecture . . . 210
B.3 User Guide . . . 211

C Presto **215**
C.1 Presto Simulator . . . 215
C.2 Presto Web . . . 217

D Moppy **221**
D.1 Introduction 221
D.2 General Idea 223
D.3 Architecture, Features, and User Interface 224

Bibliography **227**

Afterword **241**

Streszczenie **243**