

Contents

Foreword	ix
Preface	xi
I Theory and Algorithms	1
1 Theoretical Foundations	3
1.1 Basic Notions of Orders and Lattices	3
1.2 Context, Concept, and Concept Lattice	10
1.3 Many-valued Contexts	17
1.4 Bibliographic Notes	21
2 Algorithms	25
2.1 Constructing Concept Lattices	26
2.1.1 Computational space complexity of concept lattices	26
2.1.2 Construction of the set of concepts	29
2.1.3 Construction of concept lattices	34
2.1.4 Construction of partial concept lattices	39
2.2 Incremental Lattice Update	41
2.2.1 Incremental construction of concept lattices	41
2.2.2 Updating the context	52
2.2.3 Summary of lattice construction	55
2.3 Visualization	56
2.3.1 Hierarchical folders	57
2.3.2 Nested line diagrams	58
2.3.3 Focus+context views	61
2.4 Adding Knowledge to Concept Lattices	64
2.4.1 Adding background knowledge to object description	65
2.4.2 Pruning concepts with user constraints	73
2.5 Bibliographic Notes	77

II Applications	83
3 Information Retrieval	85
3.1 Query Modification	85
3.1.1 Navigating around the query concept	85
3.1.2 Thesaurus-enhanced navigation and querying	88
3.1.3 Automatic generation of index terms	93
3.2 Document Ranking	95
3.2.1 The vocabulary problem	95
3.2.2 Concept lattice-based ranking	97
3.2.3 Scalability	103
3.3 Bibliographic Notes	104
4 Text Mining	109
4.1 Mining the Content of the ACM Digital Library	109
4.1.1 The ACM Digital Library	110
4.1.2 Information retrieval and data view versus text mining	111
4.1.3 Constructing the TOIS concept lattice	112
4.1.4 Interacting with the TOIS concept lattice	120
4.2 Mining Web Retrieval Results with CREDO	127
4.2.1 Visualizing Web retrieval results	127
4.2.2 Design and implementation of CREDO	128
4.2.3 Example sessions	136
4.3 Bibliographic Notes	138
5 Rule Mining	141
5.1 Implications	141
5.1.1 Computational space complexity of implications	144
5.1.2 Generating implications from the concept lattice	147
5.2 Functional Dependencies	153
5.2.1 Functional dependencies as implications of transformed contexts	154
5.2.2 Computational space complexity of the concept lattice of transformed contexts	156
5.3 Association Rules	159
5.3.1 Mining frequent concepts	161
5.3.2 Generating confident rules from frequent concepts	164

5.4 Classification Rules	168
5.5 Bibliographic Notes	170
References	175
Index	197

Foreword

Concept lattices (*Formale Kontexte* in French) is the common name for a specialized form of Hasse diagram that is used in conceptual data processing. Concept lattices are a principled way of representing and visualizing the structure of symbolic data that emerged from Rudolf Wille's attempt to restructure (or modernize) lattice and order theory in the 1970s. Wille's objectives were to make the presentation of applied lattice and order theory – foundational work in lattice theory from the 1930s by Birkhoff and others with origins in eighteenth-century mathematics – to better promote communication between lattice theorists and its users. Conceptual data processing (also widely known as 'formal concept analysis') has become a standard technique in data and knowledge processing that has given rise to applications in data visualization, data mining, information retrieval (using ontologies) and knowledge management.

In terms of theory, formal concept analysis has been extended, enabling a general framework for knowledge representation and machine learning called "conceptual knowledge processing". The theoretical developments in the field have proceeded apace, the ideas have an impeccable intellectual pedigree and the scientific literature is vast. *Formal Concept Analysis: Mathematical Foundations* by Bernhard Ganter and Rudolf Wille (Springer-Verlag, 1999), provides a widely read reference to the basic mathematical theory, however, for computer scientists, a broad understanding of the technique and its impact has been slower to emerge. Restructuring the presentation of conceptual data processing relevant to a computer science readership is both sympathetic to Wille's restructuring intentions and is an important achievement of the present volume.

This book covers the basic theoretical and algorithmic foundations of the field and demonstrates their utility for computer scientists interested in information retrieval, data visualization and machine learning. It synthesizes a multitude of sources into a coherent pedagogical presentation that provides an implementation roadmap for practitioners and researchers interested in this important data analysis technique. It can also be used as a textbook for the study of conceptual data processing for undergraduate and postgraduate