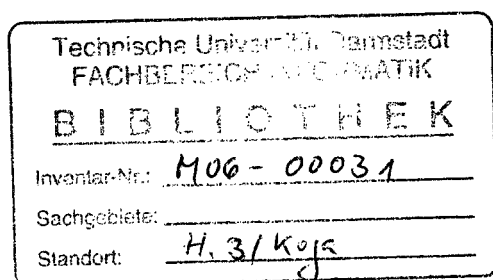


Jacob Kogan · Charles Nicholas  
Marc Teboulle (Eds.)

# Grouping Multidimensional Data

Recent Advances in Clustering

With 53 Figures



---

# Contents

|                                                                                                                                                |     |
|------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>The Star Clustering Algorithm for Information Organization</b><br><i>J.A. Aslam, E. Pelekhou, and D. Rus</i> .....                          | 1   |
| <b>A Survey of Clustering Data Mining Techniques</b><br><i>P. Berkhin</i> .....                                                                | 25  |
| <b>Similarity-Based Text Clustering: A Comparative Study</b><br><i>J. Ghosh and A. Strehl</i> .....                                            | 73  |
| <b>Clustering Very Large Data Sets<br/>with Principal Direction Divisive<br/>Partitioning</b><br><i>D. Littau and D. Boley</i> .....           | 99  |
| <b>Clustering with Entropy-Like <math>k</math>-Means Algorithms</b><br><i>M. Teboulle, P. Berkhin, I. Dhillon, Y. Guan, and J. Kogan</i> ..... | 127 |
| <b>Sampling Methods for Building Initial Partitions</b><br><i>Z. Volkovich, J. Kogan, and C. Nicholas</i> .....                                | 161 |
| <b>TMG: A MATLAB Toolbox for Generating Term-Document<br/>Matrices from Text Collections</b><br><i>D. Zeimpekis and E. Gallopoulos</i> .....   | 187 |
| <b>Criterion Functions for Clustering on High-Dimensional Data</b><br><i>Y. Zhao and G. Karypis</i> .....                                      | 211 |
| <b>References</b> .....                                                                                                                        | 239 |
| <b>Index</b> .....                                                                                                                             | 265 |