Mining Structured Data
Nijssen, Siegfried Gerardus Remius

Citation

Version: Corrected Publisher’s Version
License: Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from: https://hdl.handle.net/1887/4395

Note: To cite this publication please use the final published version (if applicable).
Bibliography


 Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD), pages 99–107. 


[192] D. Weininger. SMILES, a chemical language and information system. 1. Introduction 
 to methodology and encoding rules. In Journal of Chemical Information and Computer 

 of unique SMILES notation. In Journal of Chemical Information and Computer 

 Symposium on Graph Drawing (GD), volume 2265 of Lecture Notes in Computer 


[197] R. Wright, B. Richmond, A. Odzlyzko, and B. McKay. Constant time generation of 

 subtrees. In Proceedings of the Third IEEE International Conference on Data Mining 

[199] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining. In Proceedings of 
 the Second IEEE International Conference on Data Mining (ICDM), pages 721–724. 

[200] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining — Expanded 
 version. Technical Report UIUCDCS-R-2002-2296, University of Illinois at Urbana-
 Champaign, 2002.

[201] X. Yan and J. Han. CloseGraph: Mining closed frequent graph patterns. In Proceedings of 
 the Ninth ACM SIGKDD International Conference on Knowledge Discovery and 

[202] X. Yan, P. S. Yu, and J. Han. Graph indexing: A frequent structure-based approach. In 
 Proceedings of the ACM SIGMOD International conference on Management of Data, 

 International Conference Knowledge Discovery and Data Mining (KDD), pages 71– 


Index

\(\theta\)-subsumption, 69

AcGM, 202
ADIMine, 198
AGM, 201
Aho, Hopcroft and Ullman’s tree isomorphism algorithm, 134
Alternating path, 143
Anti-monotonicity, 40
APRIORI algorithm, 14
property, 14
APRIORISMP, 233
Association rules, 9
Augmenting path, 143
Automorphisms, 179
Backbone depth tuple order, 172
Backtracking sequence, 133
Bias, 43, 73
Bipartite involved matchings problem, 143
Bongard datasets, 92

C-ArMR, 98
CAP, 60
CBA, 237
CBAms, 237
CHARM, 64
Chemistry, 4, 94, 165, 211
CHOPPER, 151
Chung’s subtree algorithm, 141, 201
CLOSE, 63
Closed itemsets, 51
CLOSEGRAPH, 162, 197
CMAR, 237
CMTREE_MINER, 152

CN2-SD, 238
Complexities
graph isomorphism algorithms, 163
subgraph isomorphism algorithms, 163
subtree isomorphism algorithms, 110
tree isomorphism algorithms, 110
Confidence, 10
Constraint
anti-monotonic, 41
convertible, 43
monotonic, 40
Succinct, 44
Contingency table, 229
CorCLASS, 235
Cover, 10
Covers, 36
Cycle, 105

D-LAB, 97
Data tree, 110
Depth sequence, 114
Depth tuple, 114
Diffsets, 19, 191
DRYADE, 153
DualMiner, 60

ECLAT, 18
Edge sequence, 162
Enumeration, 39, 130
Equivalence classes, 35
EXAnte property, 61

FARMER, 83
FFSM, 198
Formal concept analysis (FCA), 247
Formal concepts analysis (FCA), 52
INDEX

FP-Bonsai, 61
FP-Growth, 23, 60
Free itemsets, 51
FreeTreeMiner, 200, 201
FREQ, 139
Frequent itemsets, 9
FSG, 203
FST-Forest, 151

GBI, 204
GraphML, 113, 166
Graphs, 105
Greatest lower bound, 33
GSP, 64
gSpan, 162, 196
GXL, 113, 166

HybridTreeMiner, 151, 199
Hypergraphs, 112, 165

Inductive Logic Programming (ILP), 68, 165
Itemset occurrences, 10

Kirkhoff’s matrix-tree theorem, 167
k-Prefix, 13

Large itemsets, 11
Lattice, 35
Learning from entailment, 97
Learning from interpretations, 97
Least upper bound, 33
Leftmost path, 172
Lexicographical order, 13

Maximal frequent itemsets, 50
Merge operators, 53
Merging
...of cyclic graphs, 185
...of free trees, 176
...of ordered trees, 115
...of unordered trees, 124
basic definitions, 54
downward, 55
Modes, 79
MoFA, 198
MolFea, 61

Monotonicity, 40
Multi-relational data mining, 4, 113, 165
Multicast dataset, 112

Nauty, 204
Next prefix node, 120
Non-derivable itemsets, 51

Object exchange model, 151
Object Identity, 71
Occurrence sequence, 136, 189
Occurrence tree, 145
Orders, 12

Path
rooted, 106
simple, 105

PathJoin, 151
Pattern tree, 110
PolyFarm, 99
Prefix trie, 15
PREFIXSPAN, 64
Primary key, 74
Projected database, 20
Proteins, 165
Query packs, 99

RADAR, 101
Receiver Operating Characteristic (ROC), 229
Refinement
...of cyclic graphs, 179
...of free trees, 173
...of ordered trees, 115
...of unordered trees, 117
basic definitions, 29
downward, 36
suboptimal, 31
upward, 36
Relations, 12
Rightmost path, 114

SD-Apriori, 237
Sequences, 13
SiGram, 164, 203
Simple occurrence sequence, 136
SMILES, 5
Spin, 199
Stamp point, 229
STUCCO, 235
SUBDUE, 204
Subgraphs, 161
Subpaths, 36
Subsequences
...with $(\alpha, \beta)$ gaps, 34
...with unlimited gaps, 34
...without gaps, 34
Subtrees
  bottom-up, 110
  embedded ordered, 107
  embedded unordered, 107
  induced ordered, 107
  induced unordered, 107
  ordered leaf, 109
  prefix ordered, 109
Support, 10
Symmetry, 170

Transaction, 9
Transaction based support, 47
TreeFinder, 152
TreeMinerV, 135
Trees, 105

uFreqT, 141, 148
Ullman’s subgraph isomorphism algorithm, 204
UML, 100
uNot, 140

Version space, 50
VF Algorithm, 204

WARMR, 83

XML, 113, 166
Acknowledgements

I would like to thank Eric-Wubbo Lameijer for building the molecular model of Cuneane of which a photo is included in this thesis. I enjoyed the discussions that I had with Eric-Wubbo and Jeroen Kazius about mining molecular databases. These discussions have motivated me very much, and I would like thank them for that. Of course I would also like to thank all colleagues that I used to have lunch and ‘coffee’ breaks with for making the period in Leiden an enjoyable one.