



Universiteit
Leiden
The Netherlands

Mining Structured Data

Nijssen, Siegfried Gerardus Remius

Citation

Nijssen, S. G. R. (2006, May 15). *Mining Structured Data*. Retrieved from <https://hdl.handle.net/1887/4395>

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).

Mining Structured Data

Mining Structured Data

proefschrift

ter verkrijging van
de graad van Doctor
aan de Universiteit Leiden
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op 15 mei 2006
te klokke 15.15 uur

door

Siegfried Gerardus Remius Nijssen
geboren te 's-Gravenhage
in 1978

Promotiecommissie

Promotor: Prof. Dr. J.N. Kok
Co-promotor: Dr. W.A. Kusters
Referent: Prof. Dr. L. De Raedt (Albert-Ludwigs-Universität Freiburg)
Overige leden: Prof. Dr. F. Arbab
Prof. Dr. T.H.W. Bäck
Prof. Dr. S.M. Verduyn Lunel
Prof. Dr. G. Rozenberg



The work in this thesis has been carried out under the auspices of the research school IPA (Institute for Programming research and Algorithmics).

To my parents and brothers

Contents

1	Introduction	1
1.1	Data Mining	1
1.2	Structured Data	4
1.3	Overview	6
2	Frequent Itemset Mining	9
2.1	Introduction	9
2.2	Frequent Itemset Mining Principles	11
2.3	Orders and Sequences	12
2.4	APRIORI	14
2.5	ECLAT	18
2.6	FP-GROWTH	23
2.7	Conclusions	25
3	Theory of Inductive Databases	27
3.1	Introduction	27
3.2	Searching through Space	28
3.3	Relations between Structures	33
3.4	Constraints and Inductive Queries	39
3.5	Condensed Representations	50
3.6	Mining under Monotonic Constraints; Merge Operators	53
3.7	Inductive Database Mining Algorithms	60
3.8	Frequent Sequence Mining Algorithms	64
3.9	Conclusions	65
4	Inductive Logic Databases	67
4.1	Introduction	67
4.2	First Order Logic	68
4.3	Weak Object Identity using Primary Keys	73
4.4	A Practical Refinement Operator	78
4.5	Frequent Atom Set Mining	82
4.6	FARMER	83
4.7	Depth-First and Breadth-First Algorithms	90
4.8	Experimental Results	92

4.9	Related Work	96
4.10	Conclusions	101
5	Mining Rooted Trees	103
5.1	Introduction	103
5.2	Graphs and Trees: Basic Definitions	104
5.3	Applications	112
5.4	Ordered Trees: Encodings and Refinement	114
5.5	Unordered Trees: Encodings and Refinement	117
5.6	Unordered Trees: Refinements — Proofs	125
5.7	Enumeration of Unordered Trees	130
5.8	Mining Bottom-Up Subtrees	133
5.9	Mining Induced Leaf Subtrees	135
5.10	Mining Embedded Subtrees	135
5.11	Mining Induced Subtrees using Refinement	139
5.12	Mining Unordered Induced Subtrees using Refinement: uFREQT	141
5.13	Mining Induced Subtrees using Merges	148
5.14	Related Work	151
5.15	Experimental Results	153
5.16	Conclusions	158
6	Mining Free Trees and Graphs	159
6.1	Introduction	159
6.2	Graphs and Trees: Basic Definitions — Continued	161
6.3	Applications	164
6.4	On the Complexity of Graph Mining	166
6.5	Mining Subgraphs in Uniquely Labeled Graphs	167
6.6	Paths: Encodings and Refinement	167
6.7	Free Trees: Encodings and Refinement	171
6.8	Cyclic Graphs: Encodings and Refinement	179
6.9	Evaluation using Occurrence Sequences	189
6.10	Evaluation by Recomputing Occurrences	192
6.11	Related Work	196
6.12	Experimental Results	205
6.13	Conclusions	225
7	Mining Correlated Patterns	227
7.1	Introduction	227
7.2	Plotting Frequent Patterns in ROC Space	228
7.3	Accuracy and Weighted Relative Accuracy	230
7.4	Class Neutral Measures	232
7.5	Related work	233
7.6	Higher Numbers of Classes	238
7.7	High Numbers of Classes — Proofs	241
7.8	Inductive Queries that Relate Patterns	246

7.9	Experimental Results	247
7.10	Conclusions	249
8	Conclusions	251
8.1	Summary	251
8.2	Future Work	255

