

## ALGODAN Publications 2008-2013

2013

### *Articles in refereed scientific journals*

1. P.R. Adhikari and J. Hollmén. Fast progressive training of mixture models for model selection. *Journal of Intelligent Information Systems*, 2013, pp. 1-19.
2. P. Barboza, L. Vaillant, A. Mawudeku, N.P. Nelson, D.M. Hartley, L.C. Madoff, J.P. Linge, N. Collier, J.S. Brownstein, R. Yangarber, P. Astagneau. Evaluation of Epidemic Intelligence Systems Integrated in the Early Alerting and Reporting Project for the Detection of A/H5N1 Influenza Events. *PLoS One*, vol. 8, no. 3, 2013, pp. e57252.
3. M. Bonke, M. Turunen, M. Sokolova, A. Vaharautio, T. Kivioja, M. Taipale, M. Björklund and J. Taipale. Transcriptional Networks Controlling the Cell Cycle. *G3 - Genes genomes genetics*, vol. 3, no. 1, 2013, pp. 75-90.
4. Y. Borodin, V. Polishchuk, J. Mahmud, I.V. Ramakrishnan and A. Stent. Live and learn from mistakes: A lightweight system for document classification. *Information Processing & Management*, vol. 49, no. 1, 2013, pp. 83-98.
5. R.P.J.C. Bose, W.M.P. van der Aalst, I. Zliobaite and M. Pechenizkiy. Dealing with Concept Drifts in Process Mining. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1), 2013, pp. 154-171.
6. A.G. Bunn, E. Jansma, M. Korpela, R.D. Westfall and J. Baldwin. Using simulations and data to evaluate mean sensitivity (\$\$) as a useful statistic in dendrochronology. *Dendrochronologia*, 31(3), 2013, pp. 250-254.
7. C. Campi, L. Parkkonen, R. Hari and A. Hyvärinen. Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Frontiers in Brain Imaging Methods*, vol. 7, no. 107, 2013.
8. H.G. Eyherabide and I. Samengo. When and why noise correlations are important in neural decoding. *Journal of Neuroscience*, vol. 33, no. 45, 2013, pp. 17921-17936.
9. T. Gagie, J. Kärkkäinen, G. Navarro and S.J. Puglisi. Colored range queries and document retrieval. *Theoretical Computer Science*, vol. 483, 2013, pp. 36-50.
10. E. Giaquinta and S. Grabowski. New algorithms for binary jumbled pattern matching. *Information Processing Letters*, vol. 113, no. 14-16, 2013, pp. 538-542.
11. E. Giaquinta, S. Grabowski and K. Fredriksson. Approximate pattern matching with k-mismatches in packed text. *Information Processing Letters*, vol. 113, no. 19-21, 2013, pp. 693-697.
12. E. Giaquinta, S. Grabowski and E. Ukkonen. Fast Matching of Transcription Factor Motifs Using Generalized Position Weight Matrix Models. *Journal of Computational Biology*, vol. 20, no. 9, 2013, pp. 621-630.
13. E. Giaquinta and L. Pozzi. An Effective Exact Algorithm and a New Upper Bound for the Number of Contacts in the Hydrophobic-Polar Two-Dimensional Lattice Model. *Journal of Computational Biology*, vol. 20, no. 8, 2013, pp. 593-609.
14. M.U. Gutmann and A. Hyvärinen. A three-layer model of natural image statistics. *Journal of Physiology (Paris)*, vol. 107, no. 5, 2013, pp. 369-398.

15. J. Huvelin, O. Gross, O. Solin, K. Linden, S.P.T. Maisala, T. Oittinen, H. Toivonen, J. Niemi and M. Silfverberg. Software Newsroom – an approach to automation of news search and editing. *Journal of Print Media Technology research*, vol. 2, no. 3, 2013, pp. 141-156.
16. H. Huttunen, T. Manninen, J.-P. Kauppi and J. Tohka. Mind reading with regularized multinomial logistic regression. *Machine Vision & Applications*, vol. 24, no. 6, 2013, pp. 1311-1325.
17. A. Hyttinen, F. Eberhardt and P.O. Hoyer. Experiment Selection for Causal Discovery. *Journal of Machine Learning Research*, vol. 2013, no. 14, 2013, pp. 3041-3071.
18. A. Hyvärinen and P. Ramkumar. Testing independent component patterns by inter-subject or inter-session consistency. *Frontiers in Human Neuroscience*, vol. 7, 2013.
19. A. Hyvärinen and S.M. Smith. Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models. *Journal of Machine Learning Research*, vol. 14, 2013, pp. 111-152.
20. A. Jolma, J. Yan, T. Whittington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M.J.O. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja and J. Taipale. DNA-Binding Specificities of Human Transcription Factors. *Cell*, vol. 152, no. 1-2, 2013, pp. 327-339.
21. J.-P. Kauppi, L. Parkkonen, R. Hari and A. Hyvärinen. Decoding magnetoencephalographic rhythmic activity using spectrospatial information. *NeuroImage*, vol. 83, 2013, pp. 921-936.
22. L. Langohr, V. Podpečan, M. Petek, I. Mozetič, K. Gruden, N. Lavrač and H. Toivonen. Contrasting Subgroup Discovery. *Computer Journal*, vol. 56, no. 3, 2013, pp. 289-303.
23. J. Li, J. Liu, H. Toivonen and J. Yong. Effective Pruning for the Discovery of Conditional Functional Dependencies' *Computer Journal*, vol. 56, no. 3, 2013, pp. 378-392.
24. J.M. Leppilähti, M.A. Kallio, T. Tervahartiala, T. Sorsa and P. Mantyla. Gingival Crevicular Fluid (GCF) Matrix Metalloproteinase-8 Levels Predict Treatment Outcome Among Smoking Chronic Periodontitis Patients. *Journal of Periodontology*, 85, 2013.
25. A. Moneta, D. Entner, P.O. Hoyer and A. Coad. Causal Inference by Independent Component Analysis: Theory and Applications. *Oxford Bulletin of Economics and Statistics*, vol. 75, no. 5, 2013, pp. 705-730.
26. E. Packer, P. Bak, M. Nikkilä, V. Polishchuk and H.-J. Ship. Visual Analytics for Spatial Clustering: Using a Heuristic Approach for Guided Exploration. *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2179-2188.
27. P. Parviainen and M. Koivisto. Finding Optimal Bayesian Networks Using Precedence Constraints. *Journal of Machine Learning Research*, vol. 14, 2013, pp. 1387-1415.
28. P. Rastas, L. Paulin, I. Hanski, R.J. Lehtonen and P. Auvinen. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics*, vol. 29, no. 24, 2013, pp. 3128-3134.
29. J. Rousu, D.D. Agranoff, O. Sodeinde, J. Shawe-Taylor and D. Fernandez-Reyes. Biomarker Discovery by Sparse Canonical Correlation Analysis of Complex Clinical Phenotypes of Tuberculosis and Malaria. *PLOS Computational Biology*, 9(4), 2013, 1003018.
30. H. Sasaki, M.U. Gutmann, H. Shouno and A. Hyvärinen. Correlated topographic analysis: estimating an ordering of correlated components' *Machine Learning*, vol. 92, no. 2-3, 2013, pp. 285-317.
31. H. Shen, N. Zamboni, M. Heinonen and J. Rousu. Metabolite Identification through Machine Learning — Tackling CASMI Challenge Using FingerID. *Metabolites*, 3(2), 484-505.
32. J. Yan, M. Enge, T. Whittington, K. Dave, J. Liu, I. Sur, B. Schmierer, A. Jolma, T. Kivioja, M. Taipale and J. Taipale. Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites. *Cell*, vol. 154, no. 4, 2013, pp. 801-813.

1. P.R. Adhikari and J. Hollmén. Mixture Models from Multiresolution 0-1 Data. In J. Fürnkranz, E. Üllmermeier and T. Higuchi (editors), *Proceedings of Sixteenth International Conference on Discovery Science (DS 2013)*, 2013, Springer Berlin Heidelberg, pp. 1-16.
2. M. Atkinson, M. Du, J. Piskorski, H. Tanev, R. Yangarber and V. Zavarella. Techniques for Multilingual Security-related Event Extraction from Online News. In A. Przepiórkowski et al.(editors), *Computational Linguistics—Applications, Studies in Computational Intelligence*, vol. 458, 2013, Springer-Verlag, pp. 163-186.
3. P. Austrin, P. Kaski, M. Koivisto and J. Määtä. Space–Time Tradeoffs for Subset Sum: An Improved Worst Case Algorithm. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP 2013), Part I*, Riga, Latvia, July, 2013, Lecture Notes in Computer Science, vol. 7965, pp. 45-56.
4. D. Belazzougui, F. Cunial, J. Kärkkäinen and V. Mäkinen. Versatile succinct representations of the bidirectional Burrows-Wheeler transform. In *Proceedings of the 21st Annual European Symposium on Algorithms (ESA 2013)*, Sophia Antipolis, France, September 2-4, 2013, Lecture Notes in Computer Science, vol. 8125, pp. 133-144.
5. D. Belazzougui, T. Gagie and G. Navarro. Better Space Bounds for Parameterized Range Majority and Minority. In *Proceedings of 13th International Symposium on Algorithms and Data Structures (WADS 2013)*, London, ON, Canada, August 12-14, 2013, Lecture Notes in Computer Science, vol. 8037, pp. 121-132.
6. A. Bifet, J. Read, I. Zliobaite, B. Pfahringer and G. Holmes. Pitfalls in benchmarking data stream classification and how to avoid them. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2013, pp. 465-479.
7. A. Bifet, J. Read, B. Pfahringer, G. Holmes and I. Zliobaite. CD-MOA: Change Detection Framework for Massive Online Analysis. In *Proceedings of the 20th International Symposium on Intelligent Data Analysis (IDA 2013)*, 2013, pp. 92-103.
8. F. Cicalese, T. Gagie, E. Giaquinta, E.S. Laber, Z. Lipták, R. Rizzi and A. Tomescu. Indexes for Jumbled Pattern Matching in Strings, Trees and Graphs. In *Proceedings of the 20th International Symposium on String Processing and Information Retrieval (SPIRE 2013)*, Jerusalem, Israel, October 7-9, 2013, Lecture Notes in Computer Science, vol. 8214, pp. 56-63.
9. M. Crochemore, R. Grossi, J. Kärkkäinen and G.M. Landau. A Constant-Space Comparison-Based Algorithm for Computing the Burrows-Wheeler Transform. In *Proceedings of the 24th Annual Symposium Combinatorial Pattern Matching (CPM 2013)*, Bad Herrenalb, Germany, June 17-19, 2013, Lecture Notes in Computer Science, vol. 7922, pp. 74-82.
10. D. Dolev, J.H. Korhonen, C. Lenzen, J. Suomela and J. Rybicki. Synchronous counting and computational algorithm design. In *Proceedings of the 15th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2013)*, Osaka, Japan, November 13–16, 2013, Lecture Notes in Computer Science, vol. 8255, pp. 237-250.
11. M. Du, J. Kangasharju, O. Karkulahti, L. Pivovarova and R. Yangarber. Combined analysis of news and Twitter messages. In *Proceedings of the Joint Workshop on NLP&LOD and SWAIE: SemanticWeb, Linked Open Data and Information Extraction*, 2013, pp. 41-48.
12. A. Efrat, M. Nikkilä and V. Polishchuk. Sweeping a Terrain by Collaborative Aerial Vehicles. In *Proceedings of the 21<sup>st</sup> ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '13)*, 2013, pp. 4-13.

13. D. Entner, P.O. Hoyer and P. Spirtes. Data-driven covariate selection for non-parametric estimation of causal effects. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2013, pp. 256-264.
14. F. Fomin, J.H. Korhonen and P. Golovach. On the Parameterized Complexity of Cutting a Few Vertices from a Graph. In *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science 2013 (MFCS 2013)*, Klosterneuburg, Austria, August 26-30, 2013, Lecture Notes in Computer Science, vol. 8087, pp. 421-432.
15. T. Gagie. On the Value of Multiple Read/Write Streams for Data Compression. In H. Aydinian, F. Cicalese and C. Deppe (editors), *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, 2013, Lecture Notes in Computer Science, vol. 7777, pp. 284-297.
16. T. Gagie, P. Gawrychowski and Y. Nekrich. Heaviest Induced Ancestors and Longest Common Substrings. In *Proceedings of CCCG '13*, 2013.
17. T. Gagie, D. Hermelin, G.M. Landau and O. Weimann. Binary Jumbled Pattern Matching on Trees and Tree-Like Structures. In *Proceedings of the 21st Annual European Symposium on Algorithms (ESA 2013)*, Sophia Antipolis, France, September 2-4, 2013, Lecture Notes in Computer Science, vol. 8125, pp. 517-528.
18. T. Gagie, W.-H. Hon and T.-H. Ku. New Algorithms for Position Heaps. In *Proceedings of the 24th Annual Symposium on Combinatorial Pattern Matching (CPM 2013)*, Bad Herrenalb, Germany, June 17-19, 2013, Lecture Notes in Computer Science, vol. 7922, pp. 95-106.
19. T. Gagie, K. Karhu, G. Navarro, S. Puglisi and J. Siren. Document Listing on Repetitive Collections. In *Proceedings of the 24th Annual Symposium on Combinatorial Pattern Matching (CPM 2013)*, Bad Herrenalb, Germany, June 17-19, 2013, Lecture Notes in Computer Science, vol. 7922, pp. 107-119.
20. E. Giaquinta, K. Fredriksson, S. Grabowski and E. Ukkonen. Motif Matching Using Gapped Patterns. In *Revised Selected Papers of the 24th International Workshop on Combinatorial Algorithms (IWOCA 2013)*, Rouen, France, July 10-12, 2013, Lecture Notes in Computer Science, vol. 8288, pp. 448-452.
21. O. Gross, A. Doucet and H. Toivonen. Named Entity Filtering Based on Concept Association Graphs. In *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, 2013.
22. S. Huttunen, A. Vihavainen, M. Du and R. Yangarber. Predicting Relevance of Event Extraction for the End User. In T. Poibeau, H. Saggion, J. Piskorski and R. Yangarber (editors), *Multi-source, Multilingual Information Extraction and Summarization, Theory and Applications of Natural Language Processing*, 2013, Springer-Verlag, Berlin, Heidelberg, pp. 163-176.
23. A. Hyttinen, P. Hoyer, F. Ederhardt and M. Jarvisalo. Discovering Cyclic Causal Models with Latent Variables: A General SAT-Based Procedure. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 2013, pp. 301-310.
24. D. Ienco, A. Bifet, I. Zliobaite and B. Pfahringer. Clustering Based Active Learning for Evolving Data Streams. In *Proceedings of the 16th International Conference on Discovery Science*, 2013, pp. 79-93.
25. V. Jain and E. Galbrun. Topical organization of user comments and application to content recommendation. In *Proceedings of the 22nd International World Wide Web Conference (WWW '13)*, Companion Volume, 2013, pp. 61-62.
26. M. Korotev and L. Pivovarova. Не до X": алгоритм выявления устойчивых параметров в сочетаниях слов. In *Proceedings of the International Scientific Conference on Corpus Linguistics*, 2013.

27. M. Kopotev, L. Pivovarova, N. Kochetkova and R. Yangarber. Automatic detection of stable grammatical features in n-grams. In *Proceedings of the 9th NAACL Workshop on Multiword Expressions (MWE 2013)*, 2013, Atlanta, United States.
28. J.H. Korhonen and P. Parviainen. Exact Learning of Bounded Tree-width Bayesian Networks. In *JMLR: Workshop and Conference Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2013)*, 2013, pp. 370–378.
29. A. Kotsifakos, P. Papapetrou and V. Athitsos. IBSM: Interval-Based Sequence Matching. In *SDM*, 2013.
30. J. Kärkkäinen, D. Kempa and S. Puglisi. Crochemore’s String Matching Algorithm: Simplification, Extensions, Applications. In *Proceedings of the Prague Stringology Conference 2013*, 2013, pp. 168-175.
31. J. Kärkkäinen, D. Kempa and S. Puglisi. Lightweight Lempel-Ziv Parsing. In *Proceedings of the 12th International Symposium on Experimental Algorithms (SEA 2013)*, Rome, Italy, June 5-7, 2013, Lecture Notes in Computer Science, vol. 7933, pp. 139-150.
32. J. Kärkkäinen, D. Kempa and S. Puglisi. Linear Time Lempel-Ziv Factorization: Simple, Fast, Small. In *Proceedings of the 24th Annual Symposium on Combinatorial Pattern Matching (CPM 2013)*, Bad Herrenalb, Germany, June 17-19, 2013, Lecture Notes in Computer Science, vol. 7922, pp. 189-200.
33. J. Kärkkäinen and G. Tischler. Near in Place Linear Time Minimum Redundancy Coding. In *Proceedings of the Data Compression Conference (DCC)*, 2013, pp. 411-420.
34. A. Laaksonen. Efficient and simple algorithms for time-scaled and time-warped music search. In *Proceedings of the 10<sup>th</sup> International Symposium on Computer Music Multidisciplinary Research*, 2013, pp. 621-630.
35. A. Laaksonen. Semi-automatic melody extraction using note onset time and pitch information from users. In *Proceedings of the Sound and Music Computing Conference 2013*, 2013, pp. 689-694.
36. A. Laaksonen and K. Lemström. On finding symbolic themes directly from audio using dynamic programming. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, Curitiba, Brazil, 4-8 November, 2013, pp. 47-52.
37. J. Lijffijt. A fast and simple method for mining subsequences with surprising event counts. In H. Blockeel, K. Kersting, S. Nijssen and F. Zelezný (editors), *Machine Learning and Knowledge Discovery in Databases*, 2013, Springer Berlin Heidelberg, pp. 385-400.
38. S.M. Linkola, L. Rasku and T. Ahonen. Expandable String Representation for Music Features Using Adaptive-Size Alphabets. In *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research: Sound, Music & Motion*, 2013, pp. 920-927.
39. K. Mäki-Reinikka, J. Tornainen, A. Alafuzoff, H. Kotkanen and J.M. Toivanen. *Using Galvanic Vestibular Stimulation to Sense Abstract Data*. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE 2013)*, Enschede, Netherlands, Netherlands, 12-15 November, 2013.
40. T.M. Niinimäki and M. Koivisto. Annealed Importance Sampling for Structure Learning in Bayesian Networks. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, 2013, pp. 1579-1585.
41. T.M. Niinimäki and M. Koivisto. Treedy: A Heuristic for Counting and Sampling Subset. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI-13)*, 2013, pp. 469-477.

42. J. Nouri, L. Pivovarova and R. Yangarber. MDL-based Models for Transliteration Generation. In *Proceedings of International Conference on Statistical Language and Speech Processing (SLSP 2013)*, Tarragona, Spain, 29-31 July, 2013.
43. L. Pivovarova, M. Du and R. Yangarber. Adapting the PULS Event Extraction Framework to Analyze Russian Text. In *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing (ACL 2013)*, 2013, pp. 100-109.
44. L. Pivovarova, S. Huttunen and R. Yangarber. Event representation across genre. In *Proceedings of NAACL Workshop on EVENTS: Definition, Detection, Co-reference, and Representation*, Atlanta, GA, United States, 2013.
45. S. Puglisi and D. Kempa. Lempel-Ziv factorization: Simple, fast, practical. In *Proceedings of the Meeting on Algorithm Engineering and Experiments (ALENEX)*, 2013, pp. 103-112.
46. H. Su and J. Rousu. Multilabel Classification through Random Graph Ensembles. In *Proceedings of the 5th Asian Conference on Machine Learning (ACML2013)*, 2013, pp. 404-418.
47. N. Tatti. Itemsets for Real-valued Datasets. In *Proceedings of the 13th IEEE International Conference on Data Mining (ICDM 2013)*, 2013.
48. N. Tatti and A. Gionis. Discovering Nested Communities. In *Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2013)*, 2013, pp. 32-47.
49. J. Toivanen, M. Järvisalo and H. Toivonen. Harnessing Constraint Programming for Poetry Composition. In *Proceedings of the Fourth International Conference on Computational Creativity*, 2013, pp. 160-167.
50. J. Toivanen, H. Toivonen and A. Valitutti. Automatical Composition of Lyrical Songs. In *Proceedings of the Fourth International Conference on Computational Creativity*, 2013.
51. J. Tuimala and A. Kallio. R, Programming Language. In *Encyclopedia of Systems Biology*, 2013, pp. 1809-1811.
52. A. Valitutti, A. Doucet, J. Toivanen and H. Toivonen. "Let Everything Turn Well in Your Wife": Generation of Adult Humor Using Lexical Constraints. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), Volume 2: Short Papers*, 2013, pp. 243.
53. J. Wettig, J. Nouri, K. Reshetnikov and R. Yangarber. Information-theoretic modeling of etymological sound change. In L. Borin and A. Saxena (editors), *Approaches to measuring linguistic differences, Trends in Linguistics*, no. 265, 2013, Mouton de Gruyter, pp. 507-532.
54. I. Zliobaite and J. Hollmén. Fault tolerant regression for sensor data. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2013, pp. 449-464.

#### *Technical reports and other publications*

1. D.M. Hartley, N.P. Nelson, R.R. Arthur, P. Barboza, N. Collier, N. Lightfoot, J.P. Linge, E. van der Goot, A. Mawudeku, L.C. Madoff, L. Vaillant, R. Walters, R. Yangarber, J. Mantero, C.D. Corley and J.S. Brownstein. An overview of Internet biosurveillance. In *Clinical Microbiology and Infection*, vol. 19, no. 11, 2013, pp. 1006-1013.
2. A. Hyvärinen. Independent component analysis: recent advances. In *Philosophical transactions - Royal Society, Mathematical, Physical and engineering sciences*, vol. 371, no. 1984, 2013
3. M. Shubin. Decomposing the Bacterial Phenotypic Time-series into biologically-meaningful components. In *Proceedings of the 11th Bioinformatics Research and Education Workshop*, Berlin, Germany, 3-4 May, 2013.

4. H. Toivonen, O. Gross, J.M. Toivanen and A. Valitutti. On Creative Uses of Word Associations. In *Advances in Intelligent Systems and Computing: Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, 2013, pp. 17-24.

#### Artistic work

1. O. Gross (editor) and H. Toivonen. Organization of Computational Creativity Club. 2013.
2. O. Holmqvist, D. Murphy and J. Paalasmaa. Biomusic performance, Gallery Jade, The Night of the Arts, 2013.
3. O. Holmqvist, D. Murphy and J. Paalasmaa. Biomusic performance, the 8th World Conference of Science Journalists, 2013.
4. S. Lääne, O. Gross, J. Toivanen and H. Toivonen. ArNePo: Arts, News & Poetry. 2013.
5. K. Mäki-Reinikka, J. Toivanen, A. Alafuzoff, H. Kotkanen and J. Torniainen. Brains on Art: Brain Poetry. 2013.
6. K. Mäki-Reinikka, J. Toivanen, A. Alafuzoff, H. Kotkanen and J. Torniainen. Brains on Art: The Suit. 2013.
7. K. Mäki-Reinikka, A. Alafuzoff, H. Kotkanen, J. Toivanen and J. Torniainen. Brains on Art: It's not just in your head. 2013.
8. H. Paakkanen and J. Toivanen. Oodimobiili. 2013.
9. H. Toivonen and T. Sivula. Soul Music: Making music of your dreams. 2013.
10. H. Toivonen and T. Sivula. Musical Chair. 2013.

## 2012

#### Articles in refereed scientific journals

1. A. Björklund, T. Husfeldt, P. Kaski and M. Koivisto. The traveling salesman problem in bounded degree graphs. *ACM Transactions on Algorithms*, 8 (2), Article 18, 2012, pp. 1-13.
2. F. Claude, G. Navarro, H. Peltola, L. Salmela and J. Tarhio. String matching with alphabet sampling. *Journal of Discrete Algorithms*, 11, 2012, pp. 37-50.
3. E. Congdon, S. Service, J. Wessman, J.K. Seppanen, S. Schönauer, J. Miettunen, H. Turunen, M. Koironen, M. Joukamaa, M.-R. Jarvelin, L. Palotie, J. Veijola, H. Mannila, T. Paunio and N.B. Freimer. Early Environment and Neurobehavioral Development Predict Adult Temperament Clusters. *PLoS One*, 7(7), 2012, Article e38065.
4. J.T. Eronen, M. Fortelius, F. Portmann, K. Puolamäki and C.M. Janis. Neogene Aridification of the Northern Hemisphere. *Geology*, 40(9), 2012, pp. 823-826.
5. L. Eronen and H. Toivonen. Biomine: Predicting links between biological entities using network models of heterogeneous database. *BMC Bioinformatics*, 13, 2012, Article 119.
6. E. Galbrun and P. Miettinen. From black and white to full color: extending redescription mining outside the Boolean world. *Statistical analysis and data mining*, 5 (4), 2012, pp. 284-303.
7. M. Gutmann and A. Hyvärinen. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research*, 13, 2012, pp. 307–361.
8. H. Heikinheimo, J.T. Eronen, A. Sennikov, C.D. Preston, E. Oikarinen, P. Uotila, H. Mannila and M. Fortelius. Convergence in the distribution patterns of Europe's plants and mammals is due to environmental forcing. *Journal of Biogeography*, 39(9), 2012, pp. 1633-1644.

9. M. Heinonen, H. Shen, N. Zamboni and J. Rousu. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics*, vol. 28, no. 18, 2012, pp. 2333-2341.
10. A. Hyttinen, F. Eberhardt and P.O. Hoyer. Learning Linear Cyclic Causal Models with Latent Variables. *Journal of Machine Learning Research*, vol. 13, 2012, pp. 3387-3439.
11. M. Kashif, S. Pietila, K. Artola, R.A.C. Jones, A.K. Tugume, V. Makinen and J.P.T. Valkonen. Detection of Viruses in Sweetpotato from Honduras and Guatemala Augmented by Deep-Sequencing of Small-RNAs. *Plant Disease*, vol. 96, no. 10, 2012, pp. 1430-1437.
12. J. Kim, J. S. B. Mitchell, V. Polishchuk, S. Yang and J. Zou. Routing Multi-Class Traffic Flows in the Plane. *Computational Geometry*, 45(3), 2012, pp. 99–114.
13. T. Kivioja, A.V. Vähärautio, K. Karlsson, A.W.M., Bonke, M. Enge, S. Linnarsson and J. Taipale. Counting absolute numbers of molecules using unique molecular identifiers. *Nature methods*, 9 (1), 2012, pp. 72–74.
14. A. Kotsifakos, P. Papapetrou, J. Hollmén, D. Gunopulos, V. Athitsos and G. Kollios. Hum-a-song: A Subsequence Matching with Gaps-Range-Tolerances Query-By-Humming System. *Proceedings of the Very Large Databases Endowment (PVLDB)*, 4(12), 2012, pp. 1930-1933.
15. I. Kostitsyna and V. Polishchuk. Simple Wriggling is Hard Unless You Are a Fat Hippo. *Theory of Computing Systems*, vol. 50, no. 1, 2012, pp. 93-110.
16. J. Lijffijt and S.Th. Gries. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 17(1), 2012, pp. 147-149.
17. L. Liu, K. Puolamäki, J.T. Eronen, M.M. Atabadi, E. Hernesniemi and M. Fortelius. Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. *Proceedings of the Royal Society B*, 279(1739), 2012, pp. 2793-2799.
18. H. Meronen, S.V. Henriksson, P. Raisanen and A. Laaksonen. Climate effects of northern hemisphere volcanic eruptions in an Earth System Model. *Atmospheric Research*, vol. 114, 2012, pp. 107-118.
19. V. Mäkinen, L. Salmela and J. Ylinen. Normalized N50 Assembly Metric using Gap-Restricted Co-Linear Chaining. *BMC Bioinformatics*, vol. 13, 2012, Article 255.
20. T. Niini, I. Scheinin, L. Lahti, S. Savola, F. Mertens, J. Hollmén, T. Böhling, A. Kivioja, K.H. Nord and S. Knuutila. Homozygous Deletions of Cadherin Genes in Chondrosarcoma — an Array CGH Study. *Cancer Genetics*, 205(11), 2012, pp. 588-593.
21. J. Pajula, J.-P. Kauppi and J. Tohka. Inter-Subject Correlation in fMRI: Method Validation against Stimulus-Model Based Analysis. *PLoS One*, vol. 7, no. 8, 2012, Article e41196.
22. P. Papapetrou, G. Benson and G. Kollios. Mining Poly-regions in DNA. *International Journal of Data Mining and Bioinformatics (IJDMB)*, 2012, INDERSCIENCE.
23. M.Á. Prada, J. Toivola, J. Kullaa and J. Hollmén. Three-way analysis of structural health monitoring data. *Neurocomputing*, 80, 2012, pp. 119–128.
24. P. Ramkumar, L. Parkkonen, R. Hari and A. Hyvärinen. Characterization of neuromagnetic brain rhythms over time scales of minutes using spatial independent component analysis. *Human Brain Mapping*, vol. 33, no. 7, 2012, pp. 1648-1662.
25. L. Salmela. Average complexity of backward q-gram string matching algorithms. *Information Processing Letters*, 112(11), 2012, pp. 433–437.
26. G. Sambuceti, M. Brignone, C. Marini, M. Massollo, F. Fiz, S. Morbelli, A. Buschiazzo, C. Campi, R. Piva, A.M. Massone, M. Piana and F. Frassoni. Estimating the whole bone-marrow asset in humans



- by a computational approach to integrated PET/CT imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 39, no. 8, 2012, pp. 1326-1338.
27. O. Solin, E. Ukkonen and L. Haikala. Mining the UKIDSS GPS: star formation and embedded clusters. *Astronomy & Astrophysics*, vol. 542, no 1, 2012, Article A3.
  28. N. Välimäki, S. Ladra and V. Mäkinen. Approximate all-pairs suffix/prefix overlaps. *Information and Computation*, vol. 213, 2012, pp. 49-58.
  29. J. Wessman, S. Schönauer, J. Miettunen, H. Turunen, P. Parviainen, J.K. Seppänen, E. Congdon, S. Service, M. Koironen, J. Ekelund, J. Laitinen, A. Taanila, T. Tammelin, M. Hintsanen, L. Pulkki-Råback, L. Keltikangas-Järvinen, J. Viikari, O.T. Raitakari, M. Joukamaa, M.-R. Järvelin, N. Freimer, L. Peltonen, J. Veijola, H. Mannila and T. Paunio. Temperament clusters in a normal population: implications for health and disease. *PLoS ONE*, vol. 7, no. 7, 2012, Article e33088.

#### *Refereed conference articles and articles in edited books*

1. P.R. Adhikari and J. Hollmén. Fast Progressive Training of Mixture Models for Model Selection. In J.-G. Ganascia, P. Lenca and J.-M. Petit (editors), *Proceedings of Fifteenth International Conference on Discovery Science (DS 2012)*, 2012, Springer-Verlag, pp. 194-208.
2. T. Ahonen. Compression-based Similarity Measuring in Music Information Retrieval. In *Proceedings of the Federated Computer Science Event 2012*, Helsinki, Finland, May 28-29, 2012, University of Helsinki, Department of Computer Science, Series of Publications B, Report B-2012-1, Helsinki, 2012, pp. 68-69.
3. T. E. Ahonen. Compression-based Clustering of Chromagram Data: New Method and Representations. In *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, London, UK, June, 2012, pp. 474-481.
4. E. Arkin, A. Efrat, G. Hart, I. Kostitsyna, A. Kröllner, J. S. B. Mitchell and V. Polishchuk. Scandinavian Thins on Top of Cake: on the Smallest One-Size-Fits-All Box. In *Proceedings of the Sixth International Conference on Fun with Algorithms (FUN'12)*, Venice, Italy, June, 2012, pp. 16-27.
5. A. Björklund, T. Husfeldt, P. Kaski, M. Koivisto, J. Nederlof and P. Parviainen. Fast zeta transforms for point lattices. In *Proceedings of the 23<sup>rd</sup> Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2012)*, 2012, SIAM, pp. 1436-1444.
6. D. Entner and P. Hoyer. Estimating a Causal Order among Groups of Variables in Linear Models. In *Proceedings of the 22nd International Conference on Artificial Neural Networks, Artificial Neural Networks and Machine Learning (ICANN 2012)*, Lausanne, Switzerland, September 11-14, 2012, Lecture Notes in Computer Science, vol. 7553, 2012, pp. 83-90.
7. D. Entner, P.O. Hoyer and P. Spirtes. Statistical test for consistent estimation of causal effects in linear non-Gaussian models. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS-2012)*, La Palma, Canary Islands, 2012, pp. 364-372.
8. J. Fischer, T. Gagie, T. Kopelowitz, M. Lewenstein, V. Mäkinen and N. Välimäki. Forbidden Patterns. In D. Fernández-Baca (editor), *Proceedings of 10th Latin American Symposium on Theoretical Informatics (LATIN 2012)*, Arequipa, Peru, April 16-20, 2012, Lecture Notes in Computer Science, Vol. 7256, Springer-Verlag, pp. 327-337.
9. T. Gagie, P. Gawrychowski, J. Kärkkäinen, Y. Nekrich and S.J. Puglisi. Faster Grammar-based Self-index. In *Proceedings of 6th International Conference on Language and Automata Theory and Applications (LATA 2012)*, A Coruña, Spain, March 5-9, 2012, pp. 240-251.
10. T. Gagie, K. Karhu, J. Kärkkäinen, V. Mäkinen, L. Salmela and J. Tarhio. Indexed Multi-Pattern Matching. In D. Fernández-Baca (editor), *Proceedings of 10th Latin American Symposium on*

- Theoretical Informatics (LATIN 2012)*, Arequipa, Peru, April 16-20, 2012, Lecture Notes in Computer Science, Vol. 7256, Springer-Verlag, pp. 399-407.
11. E. Galbrun and A. Kimmig. Towards Finding Relational Redescriptions. In *Proceedings of the 15th International Conference on Discovery Science (DS 2012)*, Lyon, France, October 29-31, 2012, Lecture Notes in Computer Science, vol. 7569, 2012, pp. 52-66.
  12. E. Galbrun and P. Miettinen. A Case of Visual and Interactive Data Analysis: Geospatial Redescription Mining. In *Proceedings of the ECML PKDD 2012 Workshop on Instant Interactive Data Mining*, Bristol, United Kingdom, September 24, 2012.
  13. E. Galbrun and P.A. Miettinen. Siren: An Interactive Tool for Mining and Visualizing Geospatial Redescriptions [Demo]. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*, 2012, pp. 1544-1547.
  14. S. Gaspers, M. Koivisto, M. Liedloff, S. Ordyniak and S. Szeider. On finding optimal polytrees. In *Proceedings of the 26<sup>th</sup> Conference on Artificial Intelligence (AAAI 2012)*, 2012, pp. 750-756.
  15. C.D. Giurcaneanu, P. Luosto and P. Kontkanen. On the performance of histogram-based entropy estimators. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2012)*, 2012, pp. 1-6.
  16. S. Gog, K. Karhu, J. Kärkkäinen, V. Mäkinen and N. Välimäki. Multi-Pattern Matching with Bidirectional Indexes. In *Proceedings of the 18th International Computing and Combinatorics Conference (COCOON 2012)*, Sydney, Australia, August 20-22, 2012, Lecture Notes in Computer Science, vol. 7434, 2012, pp. 384-395.
  17. O. Gross, H. Toivonen, J. Toivanen and A. Valitutti. Lexical Creativity from Word Associations. In *Proceedings of the 7th IEEE International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2012)*, 2012, pp. 35-42.
  18. M.U. Gutmann and A. Hyvärinen. Learning a selectivity-invariance-selectivity feature extraction architecture for images. In *Proceedings of International Conference on Pattern Recognition (ICPR 2012)*, Tsukuba, Japan, 2012, pp. 918-921.
  19. M. Heinonen, N. Välimäki, V. Mäkinen and J. Rousu. Efficient Path Kernels for Reaction Function Prediction. In *Proceedings of the 3rd International Conference on Bioinformatics Models, Methods and Algorithms (Bioinformatics 2012)*, Algarve, Portugal, Feb 1-4, 2012, pp. 202-207.
  20. J. Hirayama, A. Hyvärinen and S. Ishii. Structural equations and divisive normalization for energy-dependent component analysis. In *Advances in Neural Information Processing 25 (NIPS2011)*, vol. 24, Granada, Spain, 2012, pp. 1872-1880.
  21. J. Hollmén. Mixture modeling of gait patterns in sensor data. In *Proceedings of the 5th International Conference on Pervasive Technologies and Relative to Assistive Environments (PETRA 2012)*, Crete, Greece, June 6-8, 2012, ACM, Article no 48.
  22. A. Hyttinen, F. Eberhardt and P. Hoyer. Causal Discovery of Linear Cyclic Models from Multiple Experimental Data Sets with Overlapping Variables. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI'12)*, 2012, pp. 387-396.
  23. M. Järvisalo, P. Kaski, M. Koivisto and J.H. Korhonen. Finding Efficient Circuits for Ensemble Computation. In *Proceedings of the 15th International Conference on Theory and Applications of Satisfiability Testing (SAT 2012)*, Trento, Italy, June 17-20, 2012, Lecture Notes in Computer Science, vol. 7317, 2012, pp. 369-382.
  24. P. Kaski, M. Koivisto and J.H. Korhonen. Fast Monotone Summation over Disjoint Sets. In *Proceedings of the 7th International Symposium on Parameterized and Exact Computation (IPEC*

- 2012), Ljubljana, Slovenia, September 12-14, 2012, Lecture Notes in Computer Science, vol. 7535, 2012, pp. 159-170.
25. A. Kotsifakos, P. Papapetrou, J. Hollmén, D. Gunopulos and V. Athitsos. A survey of query-by-humming similarity methods. In *Proceedings of the 5th International Conference on Pervasive Technologies and Relative to Assistive Environments (PETRA 2012)*, Crete, Greece, June 6-8, 2012, ACM, Article no 5.
  26. J. Kärkkäinen, D. Kempa and S.J. Puglisi. Slashing the Time for BWT Inversion. In *Proceedings of 2012 Data Compression Conference (DCC 2012)*, IEEE Computer Society, pp. 99-108.
  27. J. Kärkkäinen, P. Mikkola and D. Kempa. Grammar Precompression Speeds Up Burrows-Wheeler Compression. In *Proceedings of the 19th International Symposium on String Processing and Information Retrieval (SPIRE 2012)*, Cartagena de Indias, Colombia, October 21-25, 2012, Lecture Notes in Computer Science, vol. 7608, 2012, pp. 330-335.
  28. L.A. Langohr and H. Toivonen. A Model for Mining Relevant and Non-redundant Information. In *Proceedings of the 27th ACM Symposium on Applied Computing (SAC 2012)*, Trento, Italy, 2012, pp. 132-137.
  29. L.A. Langohr and H. Toivonen. Retrieval of Relevant and Non-redundant Nodes. In *Proceedings of The First SDM Workshop on Dynamic Network Analysis, in conjunction with Twelfth SIAM International Conference on Data Mining*, Anaheim, California, USA, April 2012, pp. 32-40.
  30. J. Lijffijt, P. Papapetrou and K. Puolamäki. Size Matters: Finding the Most Informative Set of Window Lengths. In P.A. Flach, T. De Bie and N. Cristianini (editors), *Machine Learning and Knowledge Discovery in Databases*, 2012, Springer Berlin Heidelberg, pp. 451-466.
  31. J. Lijffijt, T. Säily and T. Nevalainen. CEECing the baseline: Lexical stability and significant change in a historical corpus. In Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen (editors), *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Research Unit for Variation, Contacts and Change in English, Helsinki, 2012.
  32. P. Luosto, C.D. Giurcaneanu and P. Kontkanen. Construction of irregular histograms by penalized maximum likelihood: a comparative study. In *Proceedings of the IEEE Information Theory Workshop (ITW'12)*, 2012, pp. 297-301.
  33. T. Niinimäki and P. Parviainen. Local Structure Discovery in Bayesian Networks. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI-12)*, 2012, pp. 634-643.
  34. J. Paalasmaa, D.J. Murphy and O. Holmqvist. Analysis of Noisy Biosignals for Musical Performance. In *Proceedings of the 11th International Symposium on Advances in Intelligent Data Analysis XI (IDA 2012)*, Helsinki, Finland, October 25-27, 2012, Lecture Notes in Computer Science, vol. 7619, 2012, pp. 241-252.
  35. J. Paalasmaa, M. Waris, H. Toivonen, L. Leppäkorpi and M. Partinen. Unobtrusive Online Monitoring of Sleep at Home. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'12)*, 2012, pp. 3784-3788.
  36. P. Papapetrou, T. Chistiakova, J. Hollmén, V. Kalogeraki and D. Gunopulos. Finding representative objects using link analysis ranking. In *Proceedings of the 5th International Conference on Pervasive Technologies and Relative to Assistive Environments (PETRA 2012)*, Crete, Greece, June 6-8, 2012, ACM, Article no 6.
  37. V. Polishchuk and T. Andersson. Socially optimal allocation of ATM resources via truthful market-based mechanisms. In *Proceedings of the 2nd SESAR Innovation Days (SID 2012)*, 2012.
  38. S. Sankararaman, K. Abu-affash, A. Efrat, S.D. Eriksson-Bique, V. Polishchuk, S. Ramasubramanian and M. Segal. Optimization Schemes for Protective Jamming. In *Proceedings of the Thirteenth ACM*

- International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2012)*, Hilton Head Island, NC, United States, 2012, pp. 65-74.
39. H. Sasaki, M.U. Gutmann, H. Shouno and A. Hyvärinen. *Topographic Analysis of Correlated Components*. In *Proceedings of the Asian Conference on Machine Learning, Journal of Machine Learning Research – Proceedings Track*, vol. 25, 2012, pp. 365-378.
  40. M.E. Skarkala, M. Maragoudakis, S. Gritzalis, L. Mitrou, H. Toivonen and P. Moen. Privacy Preservation by k-Anonymization of Weighted Social Networks. *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, Istanbul, Turkey, August 26-29, 2012, pp. 423-428.
  41. H. Su and J. Rousu. Random Graph Ensembles in Multilabel Classification. In *ICML 2012 Workshop, Object, Functional and Structured data: towards Next Generation Kernel-based Methods*, 2012.
  42. T. Tashiro, S. Shimizu, A. Hyvärinen and T. Washio. Estimation of causal orders in a linear non-gaussian acyclic model: a method robust against latent confounders. In *Proceedings of 22nd International Conference on Artificial Neural Networks: the Artificial Neural Networks and Machine Learning (ICANN 2012)*, Lausanne, Switzerland, September 11-14, 2012, Lecture Notes in Computer Science, vol. 7552, 2012, pp. 491-498.
  43. M.J. Timonen. Categorization of Very Short Documents. In *Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012)*, 2012, pp. 5-16.
  44. J. Toivanen, H. Toivonen, A. Valitutti and O. Gross. Corpus-based generation of content and form in poetry. In *Proceedings of International Conference on Computational Creativity (ICCC)*, Dublin, Ireland, May-June 2012, pp. 175-179.
  45. A. Tulilaulu, J. Paalasmaa, M. Waris and H. Toivonen. Sleep Musicalization: Automatic Music Composition from Sleep Measurements. In *Proceedings of the 11th International Symposium on Advances in Intelligent Data Analysis XI (IDA 2012)*, Helsinki, Finland, October 25-27, 2012, Lecture Notes in Computer Science, vol. 7619, 2012, pp. 392-403.
  46. T. Vartiainen and J. Lijffijt. Premodifying -ing participles in the parsed BNC. In J. Mukherjee and M. Huber (editors), *Corpus Linguistics and Variation in English: Theory and Description*, Rodopi Amsterdam, 2012, pp. 247-258.
  47. A. Valitutti. Ambiguous Lexical Resources for Computational Humor Generation. In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence (ICAART)*, Vilamoura, Algrave, Portugal, February 6-8, 2012, SciTePress, pp. 532-535.
  48. N. Välimäki. Least Random Suffix/Prefix Matches in Output-Sensitive Time. In *Proceedings of the 23rd Annual Symposium on Combinatorial Pattern Matching (CPM 2012)*, Helsinki, Finland, July 3-5, 2012, Lecture Notes in Computer Science, vol. 7354, 2012, pp. 269-279.
  49. N. Välimäki and S.J. Puglisi. Distributed String Mining for High-Throughput Sequencing Data. In *Proceedings of the 12th International Workshop on Algorithms in Bioinformatics (WABI 2012)*, Ljubljana, Slovenia, September 10-12, 2012, Lecture Notes in Computer Science, vol. 7534, 2012, pp. 441-452.
  50. J. Wettig, J. Nouri, K. Reshetnikov and R. Yangarber. Information-theoretic Methods for Analysis and Inference in Etymology. In *Proceedings of the Fifth Workshop on Information-theoretic Methods in Science and Engineering*, 2012, pp. 53-56.
  51. H. Wettig, K. Reshetnikov and R. Yangarber. Using Context and Phonetic Features in Models of Etymological Sound Change. In *Proceedings of EACL 2012 Joint: Workshop on Visualization of Linguistic Patterns and Uncovering Language History from Multilingual Resources*, Avignon, France, 2012, pp. 108-116.

## Technical reports and other publications

1. L. Eronen, P. Hintsanen and H. Toivonen. Biomine: A network-structured resource of biological entities for link prediction. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), vol. 7250, Springer-Verlag, 2012, pp. 364–378.
2. O. Gross, A. Doucet and H. Toivonen. Term Association Analysis for Named Entity Filtering. In *Proceedings of The Twenty First Text REtrieval Conference (TREC 2012)*, Gaithersburg, Maryland, November 6-9, 2012, pp. 67.
3. S.E. Heinonen, E. Koivisto, M. Heinonen and P. Heikkinen. Seismic 3D modelling: case studies from Pyhäsalmi and Kevitsa, Finland. In *Proceedings of the Seventh Symposium on the Structure, Composition and Evolution of the Lithosphere in Finland (LITHOSPHERE 2012)*, Institute of Seismology Report, no. S-56, pp. 25-28.
4. S.M. Huttunen. Tietojenkäsittelytiede: Tiedoneristäminen. V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiillilä and M. Lounela (editors), *Genreanalyysi – tekstilajitutumuksen käsikirja*. Kotimaisten kielten keskuksen julkaisuja, no. 169, Gaudeamus Helsinki University Press, Helsinki, 2012, pp. 696-706.
5. T. Hynönen, S.J. Mahler and H. Toivonen. Discovery of Novel Term Associations in a Document Collection. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 91–103.
6. A. Kimmig, E. Galbrun, H. Toivonen and L. De Raedt. Patterns and Logic for Reasoning with Networks. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 122-143.
7. J. Kärkkäinen and J. Stoye, editors. *Proceedings of the 23rd Annual Symposium on Combinatorial Pattern Matching (CPM 2012)*, Helsinki, Finland, July 3-5, 2012, Lecture Notes in Computer Science, no. 7354, Springer-Verlag, 2012.
8. L.A. Langohr, V. Podpecan, M. Petek, I. Mozetic and K. Gruden. Contrast Mining from Interesting Subgroups. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 390-406.
9. L.A. Langohr and H. Toivonen. Finding representative nodes in probabilistic graphs. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012,, pp. 218–229.
10. P. Luosto, P. Kontkanen and K. Luosto. *The normalized maximum likelihood distribution of the multinomial model class with positive maximum likelihood parameters*. Department of Computer Science, Series of Publications C, no. C-2012-5, University of Helsinki, Department of Computer Science, Helsinki, 2012.
11. A. Mozetic, N. Lavrac, V. Podpecan, P.K. Novak, H. Motaln, M. Petek, K. Gruden, H. Toivonen and K. Kulovesi. Semantic Subgroup Discovery and Cross-context Linking for Microarray Data Analysis. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 379-389.

12. J. Piskorski and R. Yangarber. Information Extraction: Past, Present and Future. In *Multi-source, Multilingual Information Extraction and Summarization: Theory and Applications of Natural Language Processing*, Springer-Verlag, 2012, pp. 23-49.
13. T. Poibeau, H. Saggion, J. Piskorski, R. Yangarber, editors. *Multi-source, Multilingual Information Extraction and Summarization. Theory and Applications of Natural Language Processing*, Springer-Verlag, 2012, Berlin, Heidelberg, 2012.
14. J. Toivanen and P. Toivanen. Konedadaa vai keskeisoruja? *Parnasso*, no. 4, 2012, pp. 22–25.
15. H. Toivonen. Network Analysis: Overview. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 144–146.
16. H. Toivonen, F. Zhou, A. Hartikainen and A.E. Hinkka. Network Compression by Node and Edge Mergers. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 199–217.
17. A. Valitutti. Creative Coding for Humor Design: A Preliminary Exploration. *Extended abstracts of the 3rd International Workshop on Computational Humor (Computational Humor 2012)*, 2012, pp. 39-40.
18. A. Valitutti, H. Toivonen, O. Gross and J. Toivanen. Decomposition and Distribution of Humorous Effect in Interactive Systems. In *Proceedings of the AAAI Fall Symposium Series on Artificial Intelligence of Humor*, AAAI Technical Reports, no. FS-12-02, 2012, pp. 96-100.
19. F. Zhou, S.J. Mahler and H. Toivonen. Review of Network Abstraction Techniques. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 166–178.
20. F. Zhou, S.J. Mahler and H. Toivonen. Simplification of Networks by Edge Pruning. In M.R. Berthold (editor), *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Vol. 7250, Springer-Verlag, 2012, pp. 179–198.

### Artistic work

1. H. Paakkanen and J. Toivanen. *Tee se kotona*. Kiitos. 2012.

## 2011

### Articles in refereed scientific journals

1. S. Angelov, S. Inenaga, T. Kivioja and V. Mäkinen. Missing pattern discovery. *Journal of Discrete Algorithms*, 9(2), 2011, pp. 153-165.
2. A. Apostolico, C. Pizzi and E. Ukkonen. Efficient algorithms for the discovery of gapped factors. *Algorithms for Molecular Biology*, 6(5), 2011.
3. E. Arkin, M. Bender, J. Mitchell and V. Polishchuk. The Snowblower Problem. *Computational Geometry*, 44(8), 2011, pp. 370-384.
4. K. Astikainen, L. Holm, E. Pitkänen, S. Szedmak and J. Rousu. Structured Output Prediction of Novel Enzyme Function with Reaction Kernels. In *Biomedical Engineering Systems and Technologies Communications of Computer and Information Science*, 127 (5), 2011, pp. 367-378.

5. A. Björklund, T. Husfeldt, P. Kaski and M. Koivisto. Covering and packing in linear space. *Information Processing Letters*, 111(21-22), 2011, pp. 1033-1036.
6. J.T. Eronen, K. Puolamäki, H. Heikinheimo, H. Lokki, A. Venäläinen, H. Mannila and M. Fortelius. The effect of scale, climate and environment on species richness and spatial distribution of Finnish birds. *Annales Zoologici Fennici*, 48(5), 2011, pp. 257-274.
7. P. Floréen, M. Hassinen, J. Kaasinen, P. Kaski, T. Musto and J. Suomela. Local approximability of max-min and min-max linear programs. *Theory of Computing Systems*, 49(4), 2011, pp. 672-697.
8. G.C. Garriga, E. Junttila and H. Mannila. Banded structure in binary matrices. *Knowledge and Information Systems*, 28(1), 2011, pp. 197-226.
9. R. Grote, J.H. Korhonen and I. Mammarella. Challenges for process-based modelling of gas exchange in mixed forests. *Forest Systems*, 20(3), 2011, pp. 389-406.
10. M. Hassinen, J. Kaasinen, E. Kranakis, V. Polishchuk, J. Suomela and A. Wiese. Analysing local algorithms in location-aware quasi-unit-disk graphs. *Discrete Applied Mathematics*, 159(15), 2011, pp. 1566-1580.
11. M. Heinonen, S. Lappalainen, T.J. Mielikäinen and J. Rousu. Computing Atom Mappings for Biochemical Reactions without Subgraph Isomorphism. *Journal of Computational Biology*, 18(1), 2011, pp. 43-58.
12. A. Hulpke, P. Kaski and P.R.J. Östergård. The number of Latin squares of order 11. *Mathematics of Computation*, 80, 2011, pp. 1197-1219.
13. A. Hyvärinen. Testing the ICA mixing matrix based on inter-subject or inter-session consistency. *NeuroImage*, 58(1), 2011, pp. 122-136.
14. A. Kallio, K. Puolamäki, M. Fortelius and H. Mannila. Correlations and co-occurrences of taxa: the role of temporal, geographic, and taxonomic restrictions. *Palaeontologia Electronica*, 14(1), 2011, pp. 4A.
15. A. Kallio, N. Vuokko, M. Ojala, N. Haiminen and H. Mannila. Randomization techniques for assessing the significance of gene periodicity results. *BMC Bioinformatics*, 12, 2011, pp. 330.
16. P. Kaski, V. Mäkinen and P.R.J. Östergård. The Cycle Switching Graph of the Steiner Triple Systems of Order 19 is Connected. *Graphs and Combinatorics*, 27(4), 2011, pp. 539-546.
17. M. Korpela, P. Nöjd, J. Hollmén, H. Mäkinen, M. Sulkava and P. Hari. Photosynthesis, temperature and radial growth of Scots pine in northern Finland: identifying the influential time intervals. *Trees - Structure and Function*, 25(2), April 2011, pp. 323-332.
18. A. Kotsifakos, P. Papapetrou, J. Hollmén and D. Gunopulos. A subsequence matching with gaps-range-tolerances framework: A query-by-humming application. *Proceedings of the VLDB Endowment*, 4(11), 2011, pp. 761-771.
19. P. Luosto and P. Kontkanen. Clustgrams: an extension to histogram densities based on the minimum description length principle. *Central European Journal of Computer Science*, 1(4), 2011, pp. 466-481.
20. T. Nevalainen, H. Raumolin-Brunberg and H. Mannila. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, 23, 2011, pp. 1-43.
21. P. Papapetrou, V. Athitsos, M. Potamias, G. Kollios and D. Gunopulos. Embedding-based Subsequence Matching in Time Series Databases. In *ACM Transactions on Database Systems (TODS)*, 36(3), 2011, pp. 17.
22. A. Pizzi, P. Rastas and E. Ukkonen. Finding significant matches of position weight matrices in linear time. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1), 2011, pp. 69-79.

23. V. Podpecan, N. Lavrac, I. Mozetic, P. Kralj Novak, I. Trajkovski, L. Langohr, K. Kulovesi, H. Toivonen, M. Petek, H. Motaln and K. Gruden. SegMine workflows for semantic microarray data analysis in Orange4WS. *BMC Bioinformatics*, 12(416), 2011.
24. A. Rizo, K. Lemström and J.M. Iñesta. Polyphonic music retrieval with classifier ensembles. *Journal of New Music Research*, 40(4), 2011, pp. 313-325.
25. L. Salmela, V. Mäkinen, N. Välimäki, J. Ylinen and E. Ukkonen. Fast scaffolding with small independent mixed integer programs. *Bioinformatics*, 27(23), 2011, pp. 3259-3265.
26. L. Salmela and J. Schröder. Correcting errors in short reads by multiple alignments. *Bioinformatics*, 27(11), 2011, pp. 1455-1461.
27. S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. Hoyer and K. Bollen. DirectLINGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research* 12, 2011.
28. Y. Sogawa, S. Shimizu, A. Hyvärinen, T. Washio, T. Shimamura and S. Imoto. Estimating Exogenous Variables in Data with More Variables than Observations. *Neural Networks*, 24(8), 2011, pp. 875-880.
29. M. Sulkava, S. Luysaert, S. Zaehle and D. Papale. Assessing and improving the representativeness of monitoring networks: The European flux tower network example. *Journal of Geophysical Research - Biogeosciences*, 116, May 2011, pp. G00J04.
30. P. Virtala, V. Berg, M.K. Kivioja, J. Purhonen, M. Salmenkivi, P. and M. Tervaniemi. The preattentive processing of major vs. minor chords in the human brain. An event-related potential study. *Neuroscience Letters*, 487(3), 2011, pp. 406-410.
31. S. Yang, J. Mitchell, J. Krozel, V. Polishchuk, J. Kim and J. Zou. Flexible Airplane Generation to Maximize Flow under Hard and Soft Constraints. *Air Traffic Control Quarterly*, 19(3), 2011, pp. 1-26.

#### *Refereed conference articles and articles in edited books*

1. P. R. Adhikari, B.B. Upadhyaya, C. Meng and J. Hollmén. Gene selection in time-series gene expression data. In M. Loog, L. Wessels, M.J.T. Reinders, and D. de Ridder (editors), *Proceedings of the 6th IAPR Conference on Pattern Recognition in Bioinformatics*, November, 2011, Lecture Notes in Bioinformatics, Vol. 7036, Springer-Verlag, pp. 145-156.
2. P. Agarwal, A. Efrat, C. Gniady, J. Mitchell, V. Polishchuk and G. Sabhnani. Distributed Localization and Clustering Using Data Correlation and the Occam's Razor Principle. In *Proceedings of 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, 2011.
3. T. Ahonen. Kolmogorov Complexity in Lyrics. In *Proceedings of AdMIRe 2011*.
4. T. Ahonen, K. Lemström and S.M. Linkola. Compression-based Similarity Measures in Symbolic, Polyphonic Music. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, Usa, October, 2011, pp. 91-96.
5. S. Alonso, M. Dominguez, M. A. Prada, M. Sulkava and J. Hollmén. Comparative analysis of power consumption in university buildings using envSOM. In J. Gama, E. Bradley and J. Hollmén (editors), *Proceedings of 10th International Symposium on Advances in Intelligent Data Analysis (IDA 2011)*, Porto, Portugal, October, 2011, Lecture Notes in Computer Science, Vol. 7014, Springer-Verlag, pp. 10-21.
6. S. Alonso, M. Sulkava, M.A. Prada, M. Dominguez, and Jaakko Hollmén. EnvSOM: a SOM algorithm conditioned on the environment for clustering and visualization. In J. Laaksonen and T. Honkela (editors), *Proceedings of 8th International Conference on Advances in Self-Organizing*



- Maps (WSOM 2011)*, Espoo, Finland, June, 2011, Aalto University, Lecture Notes in Computer Science, Vol. 6731, Springer-Verlag, pp. 61-70.
7. E. Arkin, C. Dieckmann, C. Knauer, J. Mitchell, V. Polishchuk, L. Schlipf and S. Yang. Convex Transversals. In *Proceedings of 12th International Symposium on Algorithms and Data Structures (WADS 2011)*, New York, NY, USA, August 15-17, 2011, Lecture Notes in Computer Science, Vol. 6844, Springer-Verlag, pp. 49-60.
  8. M. Atkinson, J. Piskorski, E. Van der Goot and R. Yangarber. Multilingual real-time event extraction for border security intelligence gathering. In Uffe Kock Wiil (editor), *Counterterrorism and Open Source Intelligence*, Lecture Notes in Social Networks, Vol. 2, Springer-Verlag, 2011, pp. 355-390.
  9. M.J. Brewer, M. Sulkava, H. Mäkinen, M. Korpela, P. Nöjd and J. Hollmén. Logistic fitting method for detecting onset and cessation of tree stem radius increase. In H. Yin, W. Wang and V. Rayward-Smith (editors), *Proceedings of 12th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2011)*, Norwich, UK, September 2011, Lecture Notes in Computer Science, Vol. 6936, Springer-Verlag, pp. 204-211.
  10. M. Du, P. von Etter, M. Kopotev, M. Novikov, N. Tarbeeva and R. Yangarber. Building support tools for Russian-language information extraction. In *Proceedings of Balto-Slavonic Natural Language Processing (BSNLP-2011)*, Plzeň, Czech Republic, 2011.
  11. D. Entner and P.O. Hoyer. Discovering Unconfounded Causal Relationships Using Linear Non-Gaussian Models. In *Proceedings of New Frontiers in Artificial Intelligence: JSAI-isAI 2010 Workshops*, Tokyo, Japan, November 18-19, 2010, Lecture Notes in Computer Science, Vol. 6797, Revised Selected Papers, Springer-Verlag, 2011, pp. 181-195.
  12. P. Ferragina, J. Sirén and R. Venturini. Distribution-Aware Compressed Full-Text Indexes. In *Proceedings of the 19<sup>th</sup> Annual European Symposium on Algorithms (ESA 2011)*, Saarbrücken, Germany, September, 2011, Lecture Notes in Computer Science, Vol. 6942, Springer-Verlag, pp. 760-771.
  13. E. Galbrun and P. Miettinen. From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World. In *Proceedings of SIAM International Conference on Data Mining*.
  14. M.U. Gutmann and A. Hyvärinen. Extracting coactivated features from multiple datasets. In *Proceedings of 21st International Conference on Artificial Neural Networks (ICANN 2011)*, Espoo, Finland, June 14-17, 2011, Lecture Notes in Computer Science, Vol. 6791, Springer-Verlag, pp. 323-330.
  15. M. Gutmann and J. Hirayama. Bregman divergence as general framework to estimate unnormalized statistical models. In *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*, Barcelona, Spain, 2011, pp. 283-290.
  16. S. Huttunen, A. Vihavainen, P. von Etter and R. Yangarber. Relevance prediction in information extraction using discourse and lexical features. In *Proceedings of Nordic Conference on Computational Linguistics (Nodalida-2011)*, Riga, Latvia, 2011.
  17. A. Hyttinen, F. Eberhardt and P. Hoyer. Noisy-OR Models with Latent Confounding. In *Proceedings of the twenty-seventh conference on Uncertainty in Artificial Intelligence (UAI-11)*, 2011, pp. 363-372.
  18. E. Juntila and P. Kaski. Segmented nestedness in binary data. In *Proceedings of the 11th SIAM International Conference on Data Mining (SDM'11)*, Mesa, Arizona, USA, 28-30 April, 2011, SIAM/Omnipress, pp. 235-246.

19. M. Karvonen, M. Laitinen, K. Lemström and J. Vikman. Error-Tolerant Content-Based Music-Retrieval with Mathematical Morphology. In *Proceedings of 7th International Symposium on Exploring Music Contents (CMMR 2010)*, Málaga, Spain, June 21-24, 2010, Lecture Notes in Computer Science, Vol. 6684, Revised Papers, Springer-Verlag, 2011, pp. 321-337.
20. M. Kopotev, M. Du, P. von Etter, M. Novikov, N. Tarbeeva and R. Yangarber. Building Support Tools for Russian-Language Information Extraction. In *Proceedings of 14th International Conference on Text, Speech and Dialogue (TSD 2011)*, Pilsen, Czech Republic, September 1-5, 2011, Lecture Notes in Computer Science, Vol. 6836, Springer-Verlag, pp. 380-387.
21. O. Kostakis, P. Papapetrou and J. Hollmén. ARTEMIS: Assessing the similarity of event-interval sequences. In D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis (editors), *Proceedings of the Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD 2011)*, September, 2011, Lecture Notes in Computer Science, Vol. 6912, Springer-Verlag, pp. 229-244.
22. O. Kostakis, P. Papapetrou and J. Hollmén. Distance measure for querying arrangements of temporal intervals. In *Proceedings of 4th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2011)*, Crete, Greece, May, 2011, ACM.
23. A. Kotsifakos, V. Athitsos, P. Papapetrou, J. Hollmén and D. Gunopulos. Model-based search in large time series databases. In *Proceedings of The 4th International Conference on Pervasive Technologies Related to Assistive Environment (PETRA 2011)*, Crete, Greece, May 2011, ACM.
24. J. Krozel, M. Ganji, S. Yang, J. Mitchell and V. Polishchuk. Metrics for evaluating the impact of weather on jet routes. In *Proceedings of 15th Conference on Aviation, Range, and Aerospace Meteorology*, 2011.
25. J. Krozel, S. Yang, J. Mitchell and V. Polishchuk. Strategies to Mitigate Off-Nominal Events in Super Dense Operations. In *Proceedings of AIAA Guidance, Navigation, and Control Conference*, 2011.
26. J. Kärkkäinen and T. Gagie. Counting Colours in Compressed Strings. In *Proceedings of 22nd Annual Symposium on Combinatorial Pattern Matching (CPM 2011)*, Palermo, Italy, June 27-29, 2011, Lecture Notes in Computer Science, vol. 6818, Springer-Verlag, pp. 197-207.
27. J. Kärkkäinen and S.J. Puglisi. Fixed Block Compression Boosting in FM Indexes. In *Proceedings of 18<sup>th</sup> International Symposium on String Processing and Information Retrieval (SPIRE 2011)*, Pisa, Italy, October 17-21, 2011, Lecture Notes in Computer Science, Vol. 7024, Springer-Verlag, pp. 174-184.
28. J. Kärkkäinen and S.J. Puglisi. Cache-Friendly Burrows-Wheeler Inversion. In *Proceedings of First International Conference on Data Compression, Communications and Processing (CCP 2011)*, 2011, pp. 38-42.
29. V. Laparra, M.U. Gutmann, J. Malo and A. Hyvärinen. Complex-valued independent component analysis of natural images. In *Proceedings of 21st International Conference on Artificial Neural Networks (ICANN 2011)*, Espoo, Finland, June 14-17, 2011, Lecture Notes in Computer Science, vol. 6792, Springer-Verlag, pp. 213-220.
30. M. Laitinen and K. Lemström. Dynamic Programming in Transposition and Time-Warp Invariant Polyphonic Content-Based Music Retrieval. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR'11)*, Miami, Florida, USA, October 2011, pp. 369-374.
31. K. Lemström and M. Laitinen. Transposition and time-warp invariant geometric music retrieval algorithms. In *Proceedings of 2011 IEEE International Conference on Multimedia and Expo (ICME)*, Barcelona, Spain, 2011, pp. 1-6.

32. J. Lijffijt, P. Papapetrou, K. Puolamäki and H. Mannila. Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. In *Proceedings of the European conference of Machine learning and knowledge discovery in databases - Part II*, 2011, SpringerVerlag, pp. 341-357.
33. A. Moneta, N. Chlaß, D. Entner and P.O. Hoyer. Causal Search in Structural Vector Autoregressive Models. In *Proceedings of NIPS Mini-Symposium on Causality in Time Series*, 2011, pp. 95-118.
34. T.M. Niinimäki, P. Parviainen and M. Koivisto. Partial Order MCMC for Structure Discovery in Bayesian Networks. In *Proceedings of the Twenty-Seventh Conference Conference on Uncertainty in Artificial Intelligence (UAI-11)*, 2011, AUAI Press, pp. 557-564.
35. J. Paalasmaa, L. Leppäkorpi and M. Partinen. Quantifying respiratory variation with force sensor measurements. In *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'11)*, Boston, USA, 2011, pp. 3812-3814.
36. P. Papapetrou, A. Gionis and H. Mannila. A Shapley-value Approach for Influence Attribution. In *Proceedings of the European Conference of Machine Learning and Principles and Practices of Knowledge Discovery in Databases (ECML PKDD)*, Athens, Greece, September 5-9, 2011, Lecture Notes in Computer Science, Vol. 6912, Springer-Verlag, pp. 549-564.
37. P. Parviainen and M. Koivisto. Ancestor Relations in the Presence of Unobserved Variables. In *Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2011)*, Athens, Greece, September 5-9, 2011, Lecture Notes in Computer Science, Vol. 6912, Springer-Verlag, pp. 581-596.
38. E. Pitkänen, M. Arvas and J. Rousu. Minimum mutation algorithm for gapless metabolic network evolution. In *Proceedings of International Conference of Bioinformatics Models, Methods and Algorithms (Bioinformatics 2011)*, Rome, Italy, January, 2011, pp. 28-38.
39. V. Polishchuk and M.J. Sysikaski. Faster algorithms for minimum-link paths with restricted orientations. In *Proceedings of 12th International Symposium on Algorithms and Data Structures (WADS 2011)*, New York, NY, USA, August 15-17, 2011, Lecture Notes in Computer Science, vol. 6844, Springer-Verlag, pp. 655-666.
40. J.S. Puuronen and A. Hyvärinen. Hermite Polynomials and Measures of Non-Gaussianity. In *Proceedings of 21st International Conference Artificial Neural Networks (ICANN2011)*, 2011, pp. 205-212.
41. E. Rivals, L. Salmela and J. Tarhio. Exact search algorithms for biological sequences. In M. Elloumi and A.Y. Zomaya (editors), *Algorithms in computational molecular biology: Techniques, approaches and applications, Bioinformatics: Computational Techniques and Engineering*, 2011, John Wiley & Sons, pp. 91-111.
42. J. Rousu, D. Agranoff, J. Shawe-Taylor and D. Fernandez-Reyes. Sparse Canonical Correlation Analysis for Biomarker Discovery: A Case Study in Tuberculosis. In *Proceedings of the Fifth International Workshop on Machine Learning in Systems Biology*, 2011, pp. 73-77.
43. J. Rousu and H. Su. Multi-Task Drug Bioactivity Classification with Graph Labeling Ensembles. In *Proceedings of the 6th International Conference on Pattern Recognition in Bioinformatics*, Delft, The Netherlands, November, 2011, Lecture Notes in Computer Science, Vol. 7036, Springer-Verlag, pp. 157-167.
44. H. Sasaki, M.U. Gutmann, H. Shouno and A. Hyvärinen. Learning Topographic Representations for Linearly Correlated Components. In *Proceedings of the NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

45. J. Sirén, N. Välimäki and V. Mäkinen. Indexing Finite Language Representation of Population Genotypes. In *Proceedings of 11th International Workshop on Algorithms in Bioinformatics (WABI 2011)*, Saarbrücken, Germany, September, 2011, Lecture Notes in Bioinformatics, Vol. 6833, pp. 270-281.
46. J. Toivola and J. Hollmén. Collaborative filtering for coordinated monitoring in sensor networks. In *Proceedings of the ICDMW 2011 11th IEEE International Conference on Data Mining Workshops*, Vancouver, Canada, December, 2011, IEEE Computer Society, pp. 987-994.
47. H. Toivonen, F. Zhou, A. Hartikainen and A. Hinkka. Compression of Weighted Graphs. In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, San Diego, USA, August, 2011, pp. 965-973.
48. A. Valitutti. How Many Jokes are Really Funny? Towards a New Approach to the Evaluation of Computational Humour Generators. In *Proceedings of International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2011)*, 2011, pp. 189-200.
49. N. Vuokko and P. Kaski. Significance of patterns in time series collections. In *Proceedings of the 11th SIAM International Conference on Data Mining (SDM'11)*, Mesa, Arizona, USA, 28-30 April, 2011, SIAM/Omnipress, pp. 676-686.
50. H. Wettig, S. Hiltunen and R. Yangarber. MDL-based modeling of etymological sound change in the Uralic language family. In *Proceedings of the Fourth Workshop on Information Theoretic Methods in Science and Engineering (WITMSE-2011)*, Helsinki, Finland, 2011.
51. H. Wettig, S. Hiltunen and R. Yangarber. MDL-based models for aligning etymological data. In *Proceedings of Conference on Recent Advances in Natural Language Processing (RANLP-2011)*, Hissar, Bulgaria, 2011.
52. K. Zhang and A. Hyvärinen. A general linear non-Gaussian state-space model: Identifiability, identification, and application. In *Proceedings of the Asian Conference on Machine Learning (ACML)*, 2011, pp. 113-128.

#### *Technical reports and other publications*

1. T. Ahonen, K. Lemström and S.M. Linkola. Compressing Quantized Tonal Centroid Vectors for Cover Song Identification. In *Proceedings of MIREX*, 2011.
2. T. Elomaa, J. Hollmén and H. Mannila, editors. *Discovery Science — Proceedings of the 14th International Conference (DS 2011)*, October, 2011, Lecture Notes in Computer Science, Vol. 6926, Springer-Verlag.
3. J. Gama, E. Bradley and J. Hollmén, editors. *Advances in Intelligent Data Analysis — Proceedings of the 10th International Symposium on Intelligent Data Analysis (IDA 2011)*, October, 2011, Lecture Notes in Computer Science, Vol. 7014, Springer-Verlag.
4. J. Kivinen, C. Szepesvári, E. Ukkonen and Z. Thomas, editors. *Proceedings of 22nd International Conference on Algorithmic Learning Theory*, 2011, Lecture Notes in Artificial Intelligence, Vol. 6925, Springer-Verlag.
5. L.A. Langohr, V. Podpecan, M. Petek, I. Mozetic and K. Gruden. Subgroup Discovery from Interesting Subgroups. In *Proceedings of Bioinformatics Research and Education Workshop (BREW 2011)*, Estonia, 2011.
6. V. Mäkinen. Algoritmitutkimuksen rooli bioinformatiikassa. *Tietojenkäsittelytiede*, 32, 2011, pp. 10–15.

*Articles in refereed scientific journals*

1. E. Arkin, J. Mitchell and V. Polishchuk. Maximum Thick Paths in Static and Dynamic Environments. *Computational Geometry*, 43(3), 2010, pp. 279-294.
2. M. Arvas, N.S. Haiminen, B. Smit, J. Rautio, M. Vitikainen, M. Wiebe, D. Martinez, C. Chee, J. Kunkel, C. Sanchez, M.A. Nelson, N. Pakula, M. Saloheimo, M. Penttilä and T. Kivioja. Detecting novel genes with sparse arrays. *Gene*, 467 (1-2), 2010, pp. 41-51.
3. A. Björklund, T. Husfeldt, P. Kaski and M. Koivisto. Trimmed Moebius Inversion and Graphs of Bounded Degree. *Theory of Computing Systems*, 47(3), 2010, pp. 637-654.
4. A. Björklund, T. Husfeldt, P. Kaski and M. Koivisto. Evaluation of permanents in rings and semirings. *Information Processing Letters*, 110(20), 2010, pp. 867-870.
5. C.J. Colbourn, A.D. Forbes, M.J. Grannell, T.S. Griggs, P. Kaski, P.R.J. Östergård, D.A. Pike and O. Potttonen. Properties of the Steiner triple systems of order 19. *The Electronic Journal of Combinatorics*, 17(1), 2010, pp. R98.
6. J.T. Eronen, K. Puolamäki, L. Liu, K. Lintulaakso, J. Damuth, C. Janis and M. Fortelius. Precipitation and large herbivorous mammals, Part I: Estimates from present-day communities. *Evolutionary Ecology Research*, 12(2), 2010, pp. 217-233.
7. J.T. Eronen, K. Puolamäki, L. Liu, K. Lintulaakso, J. Damuth, C. Janis and M. Fortelius. Precipitation and large herbivorous mammals, Part II: Application to fossil data. *Evolutionary Ecology Research*, 12(2), 2010, pp. 235-248.
8. P. Floréen, P. Kaski, V. Polishchuk and J. Suomela. Almost Stable Matchings by Truncating the Gale–Shapley Algorithm. *Algorithmica*, 58(1), 2010, pp. 102-118.
9. N.S. Haiminen and H. Mannila. Evaluation of BIC and cross validation for model selection on sequence segmentations. *International Journal of Data Mining and Bioinformatics*, 4(6), 2010, pp. 675-700.
10. D. Hartley, N. Nelson, R. Walters, R. Arthur, R. Yangarber, L. Madoff, J. Linge, A. Mawudeku, N. Collier, J. Brownstein, G. Thinus and N. Lightfoot. The landscape of international event-based biosurveillance. *Emerging Health Threats Journal*, 2010/3, pp. 2-18.
11. H. Heikinheimo, J.T. Eronen, A. Sennikov, C. Preston, P. Uotila, H. Mannila and M. Fortelius. Converge in distribution patterns of Europe's plants and mammals is due to environmental forcing. *Journal of Biogeography*, 2010.
12. T. Honkela, A. Hyvärinen and J.J. Väyrynen. WordICA-emergence of linguistic representations for words by independent component analysis. *Natural Language Engineering*, 16(3), 2010, pp. 277-308.
13. A. Hyvärinen, K. Zhang, S. Shimizu and P.O. Hoyer. Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity. *Journal of Machine Learning Research*, 11, 2010, pp. 1709–1731.
14. A. Hyvärinen, P. Ramkumar, L. Parkkonen and R. Hari. Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis. *NeuroImage*, 49(1), 2010, pp. 257-271.
15. A. Jolma, T. Kivioja, J. Toivonen, L. Cheng, G. Wei, M. Enge, M. Taipale, J.M. Vaquerizas, J. Yan, M.J. Sillanpää, A.W.M. Bonke, K. Palin, S. Talukder, T.R. Hughes, N.M. Luscombe, E. Ukkonen and J. Taipale. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research*, 20(6), 2010, pp. 861-873.

16. M. Korpela, H. Mäkinen, P. Nöjd, J. Hollmén and M. Sulkava. Automatic detection of onset and cessation of tree stem radius increase using dendrometer data. *Neurocomputing*, 73(10-12), June 2010, pp. 2039-2046.
17. U. Köster and A. Hyvärinen. A Two-Layer Model of Natural Stimuli Estimated with Score Matching. *Neural Computation*, 22(9), 2010, pp. 2308-2333.
18. K. Lemström, N. Mikkilä and V. Mäkinen. Filtering methods for content-based retrieval on indexed symbolic music databases. *Information retrieval*, 13(1), 2010, pp. 1-21.
19. M. Lukk, M. Kapushesky, J.T. Nikkilä, H. Parkinson, A. Goncalves, W. Huber, E. Ukkonen and A. Brazma. A global map of human gene expression. *Nature Biotechnology*, 28(4), 2010, pp. 322-324.
20. S. Luyssaert, P. Ciais, S. L. Piao, E.-D. Schulze, M. Jung, S. Zaehle, M. J. Schelhaas, M. Reichstein, G. Churkina, D. Papale, G. Abril, C. Beer, J. Grace, D. Loustau, G. Matteucci, F. Magnani, G. J. Nabuurs, H. Verbeeck, M. Sulkava, G. R. van der Werf, and I. A. Janssens. The European carbon balance. Part 3: forests. *Global Change Biology*, 16(5), May 2010, pp. 1429-1450.
21. M. Michael, F. Nicolas and E. Ukkonen. On the complexity of finding gapped motifs. *Journal of Discrete Algorithms*, 8(2), 2010, pp. 131-142.
22. M. Ojala and G. Garriga. Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 11, June 2010, pp. 1833-1863.
23. V. Mäkinen, G. Navarro, J. Sirén and N. Välimäki. Storage and Retrieval of Highly Repetitive Sequence Collections. *Journal of Computational Biology*, 17(3), 2010, pp. 281-308.
24. E. Pitkänen, J. Rousu and E. Ukkonen. Computational methods for metabolic reconstruction. *Current Opinion in Biotechnology*, 21(1), 2010, pp. 70-77.
25. J. Saarinen, E. Oikarinen, M. Fortelius and H. Mannila. The living and the fossilized: how well do unevenly distributed points capture the faunal information in a grid? *Evolutionary Ecology Research*, 12, 2010, pp. 363-376.
26. L. Salmela. Correction of sequencing errors in a mixed set of reads. *Bioinformatics*, 26(10), 2010, pp. 1284-1290.
27. T. Tanner and H. Toivonen. Predicting and preventing student failure - using the k-nearest neighbour method to predict student performance in an online course environment. *International Journal of Learning Technology*, 5(4), 2010, pp. 356-377.
28. E. Ukkonen. Geometric Point Pattern Matching in the Knuth-Morris-Pratt Way. *Journal of Universal Computer Science*, 16(14), 2010, pp. 1902-1911.
29. A. Usvasalo, E. Elonen, U.M. Saarinen-Pihkala, R. Rätty, A. Harila-Saari, P. Koistinen, E.-R. Savolainen, S. Knuutila and J. Hollmén. Prognostic classification of patients with acute lymphoblastic leukemia by using copy number profiles identified from array-based comparative genomic hybridization data. *Leukemia Research*, 34(11), November 2010, pp. 1476-1482.
30. A. Usvasalo, S. Ninomiya, R. Rätty, J. Hollmén, U.M. Saarinen-Pihkala, E. Elonen, and S. Knuutila. Focal 9p instability in hematologic neoplasias revealed by comparative genomic hybridization and single nucleotide polymorphism microarray analyses. *Genes, Chromosomes, and Cancer*, 49(4), April 2010, pp. 309-318.
31. T. Vesala, S. Launiainen, P. Kolari, J. Pumpanen, S. Sevanto, P. Hari, E. Nikinmaa, P. Kaski, H. Mannila, E. Ukkonen, S.L. Piao and P. Ciais. Autumn temperature and carbon balance of a boreal Scots pine forest in Southern Finland. *Biogeosciences*, 7(1), 2010, pp. 163-176.
32. G. Wei, G. Badis, M.F. Berger, T. Kivioja, K. Palin, M. Enge, M. Bonke, A. Jolma, M. Varjosalo, A.R. Gehrke, J. Yan, S. Talukder, M. Turunen, M. Taipale, H.G. Stunnenberg, E. Ukkonen, T.R. Hughes,

- M.L. Bulyk and J. Taipale. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO Journal*, 29(13), 2010, pp. 2147-2160.
33. L. Yetukuri, J. Tikka, J. Hollmén and M. Orevic. Functional prediction of unidentified lipids using supervised classifiers. *Metabolomics*, 6(1), April 2010, pp. 18-26.
34. H. Yu and D.P. Bertsekas. Error Bounds for Approximations from Projected Linear Equations. *Mathematics of Operations Research*, 35(2), 2010, pp. 306-329.

#### *Refereed conference articles and articles in edited books*

1. P.R. Adhikari and J. Hollmén. Patterns from multiresolution 0-1 data. In *Proceedings of the ACM SIGKDD Workshop on Useful Patterns (UP '10)*, New York, NY, USA, 2010, ACM, pp. 8-16.
2. P. R. Adhikari and J. Hollmén. Preservation of statistically significant patterns in multiresolution 0-1 data. In T. Dijkstra, E. Tsivtsivadze, E. Marchiori and T. Heskes (editors), *Pattern Recognition in Bioinformatics*, 2010, Lecture Notes in Computer Science, Springer-Verlag.
3. P. R. Adhikari and J. Hollmén. Mixture modelling of binary data. In *Statistical Mechanics of Learning and Inference*, 2010, Poster.
4. T. Ahonen. Combining Chroma Features for Cover Version Identification. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, The Netherlands, August, 2010, pp. 165-170.
5. T. Ahonen. Compressing Lists for Audio Classification. In *Proceedings of the 3rd International Workshop on Machine Learning and Music (MML 2010)*, Florence, Italy, October, 2010.
6. E. Arkin, A. Efrat, J. Mitchell, V. Polishchuk, S. Ramasubramanian, S. Sankararaman and J. Taheri. Data Transmission and Base-Station Placement for Optimizing Network Lifetime. In *DIALM-POMC '10 Proceedings of the 6th International Workshop on Foundations of Mobile Computing*, 2010, pp. 23-32.
7. D. Arroyuelo, F. Claude, S. Maneth, V. Mäkinen, G. Navarro, K. Nguyen, J.L.T. Siren and N. Välimäki. Fast In-Memory XPath Search using Compressed Indexes. In *Proceedings of the 26th IEEE International Conference on Data Engineering (ICDE 2010)*, Long Beach, USA, March, 2010, pp. 417-428.
8. K. Astikainen, E. Pitkänen, J. Rousu, L. Holm and S. Szedmak. Reaction Kernels: Structured Output Prediction Approaches for Novel Enzyme Function. In *Proceedings of the First International Conference on Bioinformatics*, 2010, pp. 48-55.
9. M. Atkinson, J. Piskorski, J. Belyaeva, S. Huttunen and R. Yangarber. Real-Time Text Mining in Multilingual News for the Creation of a Pre-frontier Intelligence Picture. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010.
10. M. Atkinson, J. Belyaeva, V. Zavarella, J. Piskorski, S. Huttunen, A. Vihavainen and R. Yangarber. News Mining for Border Security Intelligence. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2010.
11. D.P. Bertsekas and H. Yu. Distributed Asynchronous Policy Iteration in Dynamic Programming. In *Proceedings of 2010 Allerton Conference on Communication, Control, and Computing*, 2010.
12. A. Björklund, T. Husfeldt, P. Kaski and M. Koivisto. Covering and packing in linear space. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming: Part I*, Bordeaux, France, July 6-10, 2010, Lecture Notes in Computer Science, Vol. 6198, Springer-Verlag, pp. 727-737.

13. L. De Raedt, A. Kimmig, B. Gutmann, K. Kersting, V. Santos Costa and H. Toivonen. Probabilistic Inductive Querying Using ProbLog. In S. Dzeroski, B. Goethals and P. Panov (editors), *Inductive Databases and Constraint-Based Data Mining*, 2010, Springer-Verlag, pp. 229-262.
14. B. Durian, H. Peltola, L. Salmela and J. Tarhio. Bit-parallel search algorithms for long patterns. In *Proceedings of the 9th International Symposium on Experimental Algorithms*, 2010, Lecture Notes in Computer Science, Vol. 6049, Springer-Verlag, pp. 129-140.
15. F. Eberhardt, P.O. Hoyer and R. Scheines. Combining experiments to discover linear cyclic models with latent variables. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010, pp. 185-192.
16. D. Entner and P.O. Hoyer. On causal discovery from time series data using FCI. In *Proceedings of the 5th European Workshop on Probabilistic Graphical Models*, 2010, pp. 121-128.
17. P. von Etter, S. Huttunen, A. Vihavainen, M. Vuorinen and R. Yangarber. Assessment of Utility in Web Mining for the Domain of Public Health. In *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, 2010.
18. U.M. Gutmann and A. Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *JMLR Workshop and Conference Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS2010)*, 2010, pp. 297-304.
19. P. Hintsanen, H. Toivonen and P. Sevon. Fast Discovery of Reliable Subnetworks. In *Proceedings of 2010 International Conference on Advances in Social Network Analysis and Mining (ASONAM 2010)*, Odense, Denmark, 2010, pp. 104-111.
20. J. Hirayama, A. Hyvärinen and S. Ishii. Sparse and low-rank estimation of time-varying markov networks with alternating direction method of multipliers. In *Proceedings of the 17th International Conference on Neural Information Processing. Theory and Algorithms (ICONIP 2010)*, Part I, Sydney, Australia, November 22-25, 2010, pp. 371-379.
21. J. Hollmén, H. Mäkinen and P. Nöjd. Analyzing subjective expert opinions about standardization of tree-ring series. In *WorldDendro 2010 - Abstracts of the 8th International Conference on Dendrochronology*, Rovaniemi, Finland, June 2010, p. 110.
22. A. Hyttinen, F. Eberhardt and P.O. Hoyer. Causal discovery for linear cyclic models with latent variables. In *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (PGM 2010)*, pp. 153-160.
23. A. Hyvärinen. Pairwise measures of causal direction in linear non-gaussian acyclic models. In *JMLR: Workshop and Conference Proceedings of 2nd Asian Conference on Machine Learning*, pp. 1-16.
24. D. Janzing, P.O. Hoyer and B. Schölkopf. Telling cause from effect based on high-dimensional observations. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.
25. M. Karvonen, M. Laitinen, K. Lemström and J. Vikman. Applying mathematical morphology for content-based music retrieval. In *Proceedings of International Symposium on Computer Music Modeling and Retrieval*, Malaga, Spain, 21-24 June, 2010.
26. M. Kasari, H. Toivonen and P. Hintsanen. Fast Discovery of Reliable k-terminal Subgraphs. In *Proceedings of Advances in Knowledge Discovery and Data Mining: The 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Part II, 2010, Lecture Notes in Artificial Intelligence, No. 6119, pp. 168-177.
27. M. Koivisto and P. Parviainen. A space-time tradeoff for permutation problems. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2010)*, 2010, SIAM, pp. 484-492.



28. W.M. Koolen, M.K. Warmuth and J. Kivinen. Hedging Structured Concepts. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 2010)*, 2010, pp. 93-105.
29. M. Korpela, P. Nöjd, J. Hollmén, Harri Mäkinen, M. Sulkava and P. Hari. Daily temperature and daily photosynthetic production vs. Scots pine growth. In K. Mielikäinen, H. Mäkinen and M. Timonen (editors), *WorldDendro 2010 - Abstracts of the 8th International Conference on Dendrochronology*, Rovaniemi, Finland, June, 2010, pp. 261.
30. I. Kostitsyna and V. Polishchuk. Simple Wriggling is Hard unless you are a Fat Hippo. In *Fun with Algorithms*, 2010, Lecture Notes in Computer Science, Vol. 6099, pp. 272-283.
31. D.N. Krasnoshchekov, V., Polishchuk and A. Vihavainen. Shape approximation using k-order alpha-hulls. In *Proceedings of the 2010 Annual Symposium on Computational geometry (SoCG '10)*, 2010, pp. 109-110.
32. I. Krozel, J. Mitchell, V. Polishchuk and A. Pääkkö. Throughput/Complexity Tradeoffs for Routing Traffic in the Presence of Dynamic Weather. In *ICRAT 2010 Fourth International Conference on Research in Air Transportation*, 2010.
33. J. Kärkkäinen and S.J. Puglisi. Medium-Space Algorithms for Inverse BWT. In *Proceedings of 18th Annual European Symposium on Algorithms (ESA 2010)*, Part I, Liverpool, UK, September 6-8, 2010, Lecture Notes in Computer Science, Vol. 6346, Springer-Verlag, pp. 451-462.
34. G. Lejeune, A. Doucet, R. Yangarber and N. Lucas. Filtering news for epidemic surveillance: towards processing more languages with fewer resources. In *Proceedings of Fourth International Workshop on Cross Lingual Information Access at COLING 2010 (CLIA 2010)*, Beijing, China, 2010.
35. G. Lejeune, M. Hatmi, A. Doucet, S.M. Huttunen and N. Lucas. A proposal for a multilingual epidemic surveillance system. In *Revised Selected Papers of User Centric Media: First International Conference (UCMedia 2009)*, Venice, Italy, December 9-11, 2009, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, No. 17, Vol. 40, 2010, pp. 343-348.
36. K. Lemström. Towards More Robust Geometric Content-Based Music Retrieval. In *Proceedings of ISMIR 2010*, 2010, pp. 577-582.
37. K. Lemström. Transposition and time-scale invariant geometric music retrieval. In T. Elomaa, H. Mannila and P. Orponen (editors), *Algorithms and applications: Essays dedicated to Esko Ukkonen on the occasion of his 60th birthday*, Lecture Notes in Computer Science, Vol. 6060, Springer-Verlag, 2010.
38. J. Lijffijt, P. Papapetrou, and J. Hollmén. Tracking your steps on the track: Body sensor recordings of a controlled walking experiment. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA'2010)*, June, 2010, ACM, article number 58.
39. J. Lijffijt, P. Papapetrou, J. Hollmén and V. Athitsos. Benchmarking dynamic time warping for music retrieval. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA'2010)*, June, 2010, ACM, article number 59.
40. J. Linge, R. Steinberger, F. Fuat, S. Bucci, J. Belyaeva, M. Gemo, D. Al-Khudhairy, R., Yangarber and E. van der Goot. MedISys: Medical Information System. In E Asimakopoulou (ed.), *Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed FrameworksIGI Global*, 2010, pp. 131-142.
41. T. Lokki and K. Puolamäki. Canonical analysis of individual vocabulary profiling data. In *Proceedings of 2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, 2010, pp. 152-157.

42. P. Luosto, J. Kivinen and H. Mannila. Gaussian Clusters and Noise: An Approach Based on the Minimum Description Length Principle. In *Proceedings of 13th International Conference on Discovery Science (DS 2010)*, Canberra, Australia, October 6-8, 2010, Lecture Notes in Computer Science, Vol. 6332, pp. 251-265.
43. I. Mozetic, N. Lavrac, V. Podpecan, P.K. Novak, H. Motani, M. Petek, K. Gruden, H. Toivonen and K. Kulovesi. Bisociative knowledge discovery for microarray data analysis. In *Proceedings of the International Conference on Computational Creativity (ICCC-X)*, January 7-9, 2010, Lisbon, Portugal, pp. 190-199.
44. V. Mäkinen, N. Välimäki, A. Laaksonen and R. Katainen. Unified view of backward backtracking in short read mapping. In T. Elomaa, H. Mannila and P. Orponen (editors), *Algorithms and applications: Essays dedicated to Esko Ukkonen on the occasion of his 60th birthday*, Lecture Notes in Computer Science, Vol. 6060, Springer-Verlag, 2010, pp. 182-195.
45. A. Norta, R. Yangarber and L. Carlson. Utility Evaluation of Tools for Collaborative Development and Maintenance of Ontologies. In *Proceedings of 14th IEEE International Enterprise Distributed Object Computing Conference Workshops (VORTE 2010/MOST 2010)*, 2010, pp. 207-214.
46. I. Nöllenburg, V. Polishchuk and M. Sysikaski. Dynamic One-Sided Boundary Labeling. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10)*, 2010, pp. 310-319.
47. M. Ojala. Assessing data mining results on matrices with randomization. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM '10)*, December, 2010, pp. 959-964.
48. M. Ojala, G. Garriga, A. Gionis and H. Mannila. Evaluating query result significance in databases via randomizations. In *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM '10)*, April, 2010, pp. 906-917.
49. P. Parviainen and M. Koivisto. Bayesian structure discovery in Bayesian networks with less space. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, JMLR: W&CP 9, Vol. 9, 2010, pp. 589-596.
50. M. Pihlaja, M.U. Gutmann and A.J. Hyvärinen. A Family of Computationally Efficient and Simple Estimators for Unnormalized Statistical Models. In *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, 2010.
51. V. Polishchuk and A. Vihavainen. Periodic Multi-Labeling of Public Transit Lines. In *Proceedings of 6th International Conference on GIScience (GIScience 2010)*, Zurich, Switzerland, September 14-17, 2010, Lecture Notes in Computer Science, Vol. 6292, Springer-Verlag, pp. 175-188.
52. M. A. Prada, J. Hollmén, J. Toivola and J. Kullaa. Three-way Analysis of Structural Health Monitoring Data. In S. Kaski, D.J. Miller, E. Oja and A. Honkela (editors), *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010)*, August, 2010, IEEE, pp. 256-261.
53. K. Puolamäki, A. Bertone, R. Theron, O. Huisman, J. Johansson, S. Miksch, P. Papapetrou and S. Rinzivillo. Data Mining. In D. Keim, J. Kohlhammer, G. Ellis and F. Mansmann (editors), *Mastering the Information Age Solving Problems with Visual Analytics*, Chapter 4, Eurographics Association, 2010, pp. 39-56.
54. K. Puolamäki, P. Papapetrou and J. Lijffijt. Visually controllable data mining methods. In *Proceedings of IEEE International Conference on Data Mining Workshops 2010*, 2010, pp. 409-417.
55. P. Ramkumar, A. Hyvärinen, L. Parkkonen and R. Hari. Characterization of spontaneous neuromagnetic brain rhythms using independent component analysis of short-time Fourier

- transforms. In *IFMBE Proceedings of 17th International Conference on Biomagnetism Advances in Biomagnetism (Biomag 2010)*, Dubrovnik, Croatia, March 28–April 1, 2010.
56. L. Salmela and J. Tarhio. Approximate string matching with reduced alphabet. In T. Elomaa, H. Mannila and P. Orponen (editors), *Algorithms and applications: Essays dedicated to Esko Ukkonen on the occasion of his 60th birthday*, Lecture Notes in Computer Science, Vol. 6060, Springer-Verlag, 2010.
  57. J. Sirén. Sampled Longest Common Prefix Array. In *Proceedings of 21st Annual Symposium on Combinatorial Pattern Matching (CPM 2010)*, New York, USA, June 2010, Lecture Notes in Computer Science, Vol. 6129, Springer-Verlag, pp. 227-237.
  58. K. Sirvio and J. Hollmén. Multi-year network level road maintenance programming by genetic algorithms and variable neighbourhood search. In *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems*, September, 2010, pp. 581-586.
  59. Y. Sogawa, S. Shimizu, A. Hyvärinen, T. Washio, T. Shimamura and S. Imoto. Discovery of exogenous variables in data with more variables than observations. In *Proceedings of 20th International Conference on Artificial Neural Networks (ICANN 2010)*, Part I, Thessaloniki, Greece, September 15-18, 2010, Lecture Notes in Computer Science, Vol. 6352, Springer-Verlag, pp. 67-76.
  60. H. Su, M. Heinonen and J. Rousu. Multilabel Classification of Drug-like Molecules via Max-Margin Conditional Random Fields. In *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models (PGM-2010)*, 2010, pp. 265-272.
  61. H. Su, M. Heinonen and J. Rousu. Structured Output Prediction of Anti-Cancer Drug Activity. In *Pattern Recognition in Bioinformatics*, 2010, Lecture Notes in Computer Science, Vol. 6282, Springer-Verlag, pp. 38-49.
  62. M. Timonen, P. Silvonon and M. Kasari. Modelling a Query Space Using Associations. In *Proceedings of the 20th European-Japanese Conference on Information Modelling and Knowledge Bases*.
  63. J. Toivola, M.A. Prada and J. Hollmén. Novelty detection in projected spaces for structural health monitoring. In *Proceedings of the 9th International Symposium on Advances in Intelligent Data Analysis (IDA 2010)*, Tuscon, Arizona, May 2010, Lecture Notes in Computer Science, Vol. 6065, Springer-Verlag, pp. 208-219.
  64. H. Toivonen, S. Mahler and F. Zhou. A Framework for Path-Oriented Network Simplification. In *Proceedings of 9th International Symposium on Advances in Intelligent Data Analysis (IDA 2010)*, Tuscon, Arizona, May 2010, Lecture Notes in Computer Science, Vol. 6065, Springer-Verlag, pp. 220-231.
  65. A. Ukkonen. The support vector tree. In *Algorithms and Applications*, 2010, pages 244-259.
  66. A. Ukkonen and M. Arias. Example-dependent basis vector selection for kernel-based classifiers. In *Proceedings of ECML/PKDD*, 2010, pp. 338-353.
  67. N. Vuokko. Consecutive ones property and spectral ordering. In *Proceedings of the 10th SIAM International Conference on Data Mining (SDM '10)*, 2010, pp. 350-360.
  68. N. Vuokko and P. Kaski. Testing the significance of patterns in data with cluster structure. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM2010)*, Sydney, Australia, December 14—17, 2010, pp. 1097-1102.
  69. N. Välimäki, S. Ladra and V. Mäkinen. Approximate All-Pairs Suffix/Prefix Overlaps. In *Proceedings of 21st Annual Symposium on Combinatorial Pattern Matching (CPM 2010)*, New York, USA, June 2010, Lecture Notes in Computer Science, Vol. 6129, Springer-Verlag, pp. 76-87.

70. J. Wettig, S. Hiltunen and R. Yangarber. Hidden Markov Models for Induction of Morphological Structure of Natural Language. In *Proceedings of Workshop on Information Theoretic Methods in Science and Engineering (WITMSE-2010)*, Tampere, Finland, August 2010.
71. H. Yu. Convergence of Least Squares Temporal Difference Methods under General Conditions. In *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010.
72. V. Zarzoso and A.J. Hyvärinen. Iterative algorithms. In *Handbook on Independent Component Analysis and Blind Source Separation*, Academic Press, 2010.
73. K. Zhang and A. Hyvärinen. Source separation and higher-order causal analysis of MEG and EEG. In *Proceedings of the Twenty-Sixth Conference Uncertainty in Artificial Intelligence (UAI 2010)*, 2010, pp. 709-716.
74. K. Zhang and A. Hyvärinen. Distinguishing causes from effects using nonlinear acyclic causal models. In *JMLR Workshop and Conference Proceedings: Causality: Objectives and Assessment (NIPS 2008)*, Volume 6, 2010, pp. 157–164.
75. F. Zhou, S.J. Mahler and H. Toivonen. Network Simplification with Minimal Loss of Connectivity. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010)*, Sydney, Australia, December, 2010, pp. 659-668.

#### *Technical reports and other publications*

1. M. Arita, M. Heinonen and J. Rousu, editors. *Mass Spectrometry Informatics in Systems Biology: Abstracts of the Workshop*. Series of Publications C, University of Helsinki, Department of Computer Science, 2010.
2. D. Bertsekas and H. Yu. *Q-Learning and Enhanced Policy Iteration in Discounted Dynamic Programming (Revised)*. Technical Report, C-2010-10, Series of Publications C, University of Helsinki, Department of Computer Science, 2010.
3. T. Elomaa, H. Mannila and P. Orponen, editors. *Algorithms and Applications: Essays Dedicated to Esko Ukkonen on the Occasion of His 60th Birthday*, Lecture Notes in Computer Science, Vol. 6060, Springer-Verlag, 2010.
4. A. Hyvärinen. Statistical models of natural images and cortical visual representation. *Topics in Cognitive Science*, 2(2), 2010, pp. 251-264.
5. J. Lijffijt, P. Papapetrou, N. Vuokko and K. Puolamäki. *The smallest set of constraints that explains the data: a randomization approach*. Technical Report TKK-ICS-R31, Aalto University School of Science and Technology, Department of Information and Computer Science, Espoo, Finland, May 2010.
6. L. Salmela. Merkkijonoalgoritmeja monen hahmon hakuun. *Tietojenkäsittelytiede*, 31, 2010, pp. 70–83.
7. H. Toivonen. Frequent Pattern. In C. Sammut and G.I. Webb (editors), *Encyclopedia of Machine Learning*, 1st edition, Springer-Verlag, 2010.
8. H. Toivonen. Apriori Algorithm. In C. Sammut and G.I. Webb (editors), *Encyclopedia of Machine Learning*, 1st edition, Springer-Verlag, 2010.
9. H. Toivonen. Association Rule. In C. Sammut and G.I. Webb (editors), *Encyclopedia of Machine Learning*, 1st edition, Springer-Verlag, 2010.
10. H. Toivonen. Basket Analysis. In C. Sammut and G.I. Webb (editors), *Encyclopedia of Machine Learning*, 1st edition, Springer-Verlag, 2010.
11. H. Toivonen. Frequent Itemset. In C. Sammut and G.I. Webb (editors), *Encyclopedia of Machine Learning*, 1st edition, Springer-Verlag, 2010.

12. A. Ukkonen. Approximate top-k retrieval from hidden relations. *CoRR*, abs/1008.5057, 2010.
13. E. Ukkonen. Tila ja tulevaisuus: pysähtyneisyydestä uuteen vauhtiin. *Tietojenkäsittelytiede*, 30, 2010, pp. 4-6.
14. H. Yu. *Convergence of Least Squares Temporal Difference Methods under General Conditions*. Report, C-2010-1, Series of Publications C, University of Helsinki, Department of Computer Science, 2010.
15. H. Yu. *Least Squares Temporal Difference Methods: An Analysis under General Conditions*. Technical Report, C-2010-39, Series of Publications C, University of Helsinki, Department of Computer Science, 2010.

## 2009

### Books

1. Hyvärinen, Aapo; Hurri, Jarmo; Hoyer, Patrik O. *Natural Image Statistics*. New York: Springer, 2009.

### Articles in refereed scientific journals

1. Antti Ajanki, David R. Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. Can eyes reveal interest?—Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 19:307–339, 2009.
2. Ala-Härkönen, Väinö; Brunberg, Jussi; Lemström, Kjell; Mikkilä, Niko. A content-based music retrieval engine: Musicae scientiae. *The Journal of the European Society for the Cognitive Sciences of Music 4B*, pp. 235-256, 2009
3. Arkin, Esther M; Bae, Sang Won; Efrat, Alon; Okamoto, Kazuya; Mitchell, Joseph pp. B; Polishchuk, Valentin. Geometric stable roommates. *Information processing letters 109 (2009): 4*, pp. 219-224.
4. Arkin, Esther M; Fekete, Sándor P; Islam, Kamrul; Meijer, Henk; Mitchell, Joseph pp. B; Nunez-Rodriguez, Yurai; Polishchuk, Valentin; Rappaport, David. ; Xiao, Henry. Not being (super)thin or solid is hard: a study of grid Hamiltonicity. *Computational geometry: theory and applications 42 (2009): 6-7*, pp. 582-605.
5. Bertsekas, Dimitri P; Yu, Huizhen. Projected equation methods for approximate solution of large linear systems. *Journal of computational and applied mathematics 227 (2009)*, pp. 27-50.
6. Bingham, Ella; Mannila, Heikki. Complexity control in a mixture model by the Hardy-Weinberg equilibrium. *Computational statistics & data analysis 53 (2009): 5*, pp. 1711-1719.
7. Andreas Björklund, Thore Husfeldt, Mikko Koivisto. Set Partitioning via Inclusion-Exclusion. *SIAM Journal of Computing 39(2): 546-563*, 2009
8. Tomás Feder, Heikki Mannila, and Evimaria Terzi. Approximating the minimum chain completion problem. *Information Processing Letters*, 109(17):980–985, 2009.
9. Fischer, Johannes; Mäkinen, Veli; Navarro, Gonzalo. Faster entropy-bounded compressed suffix trees. *Theoretical computer science 410 (2009)*, pp. 5354-5364.
10. Henriksson, Linda; Hyvärinen, Aapo; Vanni, Simo. Representation of cross-frequency spatial phase relationships in human visual cortex. *Journal of neuroscience 29 (2009): 45*, pp. 14342-14351.
11. Jouhten, Paula; Pitkänen, Esa; Pakula, Tiina; Saloheimo, Markku; Penttilä, Merja; Maaheimo, Hannu. 13C-metabolic flux ratio and carbon path analyses confirmed that *Trichoderma reesei* uses

- primarily the respiratory pathway also on the preferred carbon source glucose. *BMC systems biology* 3 (2009): 104, 16 pp.
12. Kaski, Petteri; Östergård, Patric R. J; Pottonen, Olli; Kiviluoto, Lasse. A catalogue of the Steiner triple systems of order 19. *Bulletin of the Institute of Combinatorics and its Applications* 57 (2009), pp. 35-41.
  13. Kaski, Petteri; Östergård, Patric R. J. Classification of resolvable balanced incomplete block designs. *Mathematica Slovaca* 59 (2009), pp. 121-136.
  14. Kaski, Petteri; Östergård, Patric R. J. There are 1,132,835,421,602,062,347 nonisomorphic one-factorizations of K14. *Journal of combinatorial designs* 17 (2009), pp. 147-159.
  15. Korhonen, Janne; Martinmaki, Petri; Pizzi, Cinzia; Rastas, Pasi; Ukkonen, Esko. MOODS. *Bioinformatics* 25 (2009): 23, pp. 3181-3182.
  16. I. Kurki, T. Peromaa, A. Hyvärinen and J. Saarinen. Visual Features Underlying Perceived Brightness as Revealed by Classification Images. *PLoS ONE*, 46(13):2009-2014, 2006.
  17. Linge, J. P; Steinberger, R; Weber, T. P ; Yangarber, Roman; van der Goot, E; Khudhairy, D. H; Stilianakis, N. I. Internet surveillance systems for early alerting of health threats. *Euro surveillance European communicable disease quarterly* 14 (2009): 13, 2 pp.
  18. L. H. Liow, M. Fortelius, K. Lintulaakso, Heikki Mannila, and N. Chr. Stenseth. Lower extinction risk in sleep-or-hide mammals. *American Naturalist*, 173:264–272, 2009.
  19. Mavroeidis, Dimitrios; Bingham, Ella. Enhancing the stability and efficiency of spectral ordering with partial supervision and feature selection. *Knowledge and information systems*. 23 pp.
  20. Muhammed Miah, Gautam Das, Vagelis Hristidis, and Heikki Mannila. Determining attributes to maximize visibility of objects. *IEEE Trans. Knowl. Data Eng.*, 21(7):959–973, 2009.
  21. Nicolas, François; Mäkinen, Veli; Ukkonen, Esko. Efficient construction of maximal and minimal representations of motifs of a string. *Theoretical computer science* 410 (2009), pp. 2999-3005.
  22. Markus Ojala, Niko Vuokko, Alekski Kallio, Niina Haiminen, and Heikki Mannila. Randomization methods for assessing data analysis results on real-valued matrices. *Statistical Analysis and Data Mining*, 2(4):209–230, November 2009.
  23. Pitkänen, Esa; Jouhten, Paula; Rousu, Juho. Inferring branching pathways in genome-scale metabolic networks. *BMC systems biology* 3 (2009): 103, 22 pp.
  24. Polishchuk, Valentin; Suomela, Jukka. A simple local 3-approximation algorithm for vertex cover. *Information processing letters* 109 (2009): 12, pp. 642-645.
  25. Shimitsu, Shohei; Hoyer, Patrik O; Hyvärinen, Aapo. Estimation of linear non-Gaussian acyclic models for latent factors. *Neurocomputing* 72 (2009): 7-9, pp. 2024-2027.
  26. Tuupanen, Sari; Turunen, Mikko; Lehtonen, Rainer; Hallikas, Outi; Vanharanta, Sakari; Kivioja, Teemu; Björklund, Mikael; Wei, Gonghong; Yan, Jian; Niittymäki, Iina; Mecklin, Jukka-Pekka; Järvinen, Heikki J; Ristimäki, Ari; Di-Bernardo, Mariachiara; East, Phil; Carvajal-Carmona, Luis; Houlston, Richard S; Tomlinson, Ian; Palin, Kimmo; Ukkonen, Esko.; Karhu, Auli; Taipale, Jussi; Aaltonen, Lauri A. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nature genetics* 41 (2009): 8, pp. 885-890.
  27. Antti Ukkonen, Kai Puolamäki, Aristides Gionis, and Heikki Mannila. A randomized approximation algorithm for computing bucket orders. *Information Processing Letters*, 109(7):356–359, 2009.
  28. Ukkonen, Esko. Maximal and minimal representations of gapped and non-gapped motifs of a string. *Theoretical computer science* 410 (2009): 43, pp. 4341-4349.
  29. Vesala, Timo; Launiainen, Samuli; Kolari, P; Pumpanen, J; Sevanto, Sanna; Hari, P; Nikinmaa, E; Kaski, Petteri; Mannila, Heikki; Ukkonen, Esko; Piao, S; Ciais, P. Autumn warming and carbon

- balance of a boreal Scots pine forest in Southern Finland. *Biogeosciences discussions* 6 (2009), pp. 7053-7081.
30. Vingron, Martin; Brazma, Alvis; Coulson, Richard; van Helden, Jacques; Manke, Thomas; Palin, Kimmo; Sand, Olivier; Ukkonen, Esko. Integrating sequence, evolution and functional genomics in regulatory genomics. *Genome biology* 10 (2009): 1, 202, 8 pp.
  31. Välimäki, Niko; Mäkinen, Veli; Gerlach, W; Dixit, K. Engineering a compressed suffix tree implementation. *Journal of experimental algorithmics* 14 (2009): 4.2, 23 pp.
  32. Wessman, Jaana; Paunio, Tiina; Tuulio-Henriksson, Annamari; Koivisto, Mikko; Partonen, Timo; Suvisaari, Jaana; Turunen, Joni A; Wedenoja, Juho; Hennah, William; Pietilainen, Olli; Lonnqvist, Jouko; Mannila, Heikki; Peltonen, Leena. Mixture model clustering of phenotype features reveals evidence for association of DTNBP1 to a specific subtype of schizophrenia. *Biological psychiatry* 66 (2009) 11, s. 990-996.
  33. Michaela Wrage, Salla Ruosaari, Paul P. Eijk, Jussuf T. Kaifi, Jaakko Hollmén, Emre F. Yekebas, Jacob R. Izbicki, Ruud H. Brakenhoff, Thomas Streichert, Sabine Riethdorf, Bauke Ylstra, Klaus Pantel, and Harriet Wikman. Genomic profiles associated with early micrometastasis in lung cancer: Relevance of 4q deletion. *Clinical Cancer Research*, 15(5):1566–1574, March 2009.
  34. Yu, Huizhen; Bertsekas, Dimitri P. Convergence results for some temporal difference methods based on least squares. *IEEE transactions on automatic control* 54 (2009): 7, pp. 1515-1531.

#### *Refereed conference articles and articles in edited books*

1. Ahonen, Teppo E. Measuring harmonic similarity using PPM-based compression distance. *Proceedings of the Workshop on Exploring Musical Information Spaces, ECDL 2009*, Corfu, Greece, October 2009. - cop. 2009. pp. 52-55.
2. M Atkinson, J Piskorski, H Tanev, E van der Goot, R Yangarber, V Zavarella. Automated Event Extraction in the Domain of Border Security. *In Proceedings of MINUCS-2009: Mining User-Generated Content for Security at the UCMedia-2009: ICST Conference on User-Centric Media* (2009) Venice, Italy.
3. Björklund, Andreas; Husfeldt, Thore; Kaski, Petteri; Koivisto, Mikko. Counting paths and packings in Halves. *Algorithms - ESA 2009*, edited by Amos Fiat, Peter Sanders. pp. 578-586.
4. E. Canfora, D. Papale, S. Luyssaert, and Mika Sulkava. Does the European eddy flux tower network represent the climatic and ecophysiological diversity of Europe?. *In European Geosciences Union General Assembly 2009, volume 11 of Geophysical Research Abstracts*, Vienna, Austria, April 2009.
5. Coulson, R; Manke, T; Palin, K; Roeder, H; Sand, O; van Helden, J; Ukkonen, Esko ; Vingron, M; Brazma, A. From gene expression profiling to gene regulation. *Modern genome annotation*. pp. 105-119.
6. Floréen, Patrik; Kaasinen, Joel; Kaski, Petteri; Suomela, Jukka. An optimal local approximation algorithm for max-min linear programs. *SPAA'09*. pp. 260-269.
7. Gutmann, Michael; Hyvärinen, Aapo. Learning features by contrasting natural images with noise. *Proceedings of the 19th International Conference on Artificial Neural Networks (ICANN-09)*, Limassol, Cyprus, 14-17 September, 2009. pp. 623-632.
8. Gutmann, Michael; Hyvärinen, Aapo. Learning reconstruction and prediction of natural stimuli by a population of spiking neurons. *European 17th Symposium on Artificial Neural Networks: - 2009*. 6 pp.

9. Gutmann, Michael; Hyvärinen, Aapo. Unsupervised learning by discriminating data from artificial noise. *Proceedings of NIPS 2009 Workshop: 4 pp.*
10. Sami Hanhijärvi, Gemma C. Garriga, and Kai Puolamäki. Randomization techniques for graphs. *In Proceedings of the 2009 SIAM International Conference on Data Mining (SDM09), 2009.*
11. Sami Hanhijärvi, Markus Ojala, Niko Vuokko, Kai Puolamäki, Nikolaj Tatti, and Heikki Mannila. Tell me something I don't know: Randomization strategies for iterative data mining. *In Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD09), 2009.*
12. Hannes Heikinheimo, Jilles Vreeken, Arno Siebes, and Heikki Mannila. Low-entropy set selection. *In SDM 2009, pages 569–580, 2009.*
13. Huseyin Hakkoymaz, Georgios Chatzimilioudis, Dimitrios Gunopulos, and Heikki Mannila. Applying electromagnetic field theory concepts to clustering with constraints. *In ECML/PKDD, pages 485–500, 2009.*
14. Hoyer, Patrik O; Hyttinen, Antti. Bayesian discovery of linear acyclic causal models. *Uncertainty in Artificial Intelligence: UAI-2009. 9 pp.*
15. Hoyer, Patrik O; Janzing, Dominik; Mooij, Joris; Peters, Jonas; Scholkopf, Bernhard. Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems 21: edited by Daphne Koller . et al. - NIPS Foundation cop. 2009 pp. 689-696.*
16. Jursic, Matjaz; Lavrac, Nada; Mozetic, Igor; Podpecan, Vid; Toivonen, Hannu. Constructing information networks from textual documents. *Proceedings of the Workshop on Explorative Analytics of Information Networks 2009 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Bled, Slovenia, September 11, 2009 / - -, pp. 23-38.*
17. Koivisto, Mikko. Partitioning into sets of bounded cardinality. *Parameterized and exact computatio. pp. 258-263.*
18. Kärkkäinen, Juha; Manzini, Giovanni; Puglisi, Simon. Permuted longest-common-prefix array. *In Proceedings of the 20th Symposium on Combinatorial Pattern Matching (CPM 2009), pp. 181-192.*
19. Köster, Urs; Lindgren, Jussi T; Gutmann, Michael; Hyvärinen, Aapo. Learning natural image structure with a horizontal product model. *Independent component analysis and signal separation. pp. 507-514.*
20. Köster, Urs; Lindgren, Jussi T; Hyvärinen, Aapo. Estimating Markov random field potentials for natural images. *Independent component analysis and signal separation. pp. 515-522.*
21. Langohr, Laura; Toivonen, Hannu. Finding representative nodes in probabilistic graphs. *Proceedings of the Workshop on Explorative Analytics of Information Networks 2009, 2009. pp. 65-76.*
22. Lemström, Kjell; Wiggins, Geraint A. Formalizing invariances for content-based music retrieval. *ISMIR 2009 - The International Society for Music Information Retrieval, 2009. pp. 591-596.*
23. Heikki Mannila and Evimaria Terzi. Finding links and initiators: A graph-reconstruction problem. *In SDM 2009, pages 1207–1217, 2009.*
24. Mäkinen, Veli; Navarro, Gonzalo; Sirén, Jouni; Välimäki, Niko. Storage and retrieval of individual genomes. *Research in computational molecular biology. pp. 121-137.*
25. Pekka Nöjd, Mikko Korpela, Jaakko Hollmén, H. Mäkinen, Mika Sulkava, and P. Hari. Effects of daily temperature and photosynthetic production on growth variation of Scots pine in northern Finland. *In Eurodendro 2009 – European Workshop in Dendrochronology: Developments, Advances, Challenges, page 85, Cala Millor, Spain, October 2009.*



26. Markus Ojala and Gemma Garriga. Permutation tests for studying classifier performance. *In Proceedings of the 9th IEEE International Conference on Data Mining (ICDM '09)*, pages 908–913, December 2009.
27. Parviainen, Pekka; Koivisto, Mikko. Exact structure discovery in Bayesian networks with less space. *Uncertainty in Artificial Intelligence: UAI-2009*. pp. 436-443.
28. Polishchuk, Valentin; Pääkkö, Anne; Mitchell, Joseph pp. B; Krozel, Jimmy. Planning routes with wiggle room in en route weather-impacted airspaces. *Proceedings of AIAA Guidance, Navigation, and Control Conference*, 10-13 August 2009, Chigaco, IL, USA.
29. Kai Puolamäki and Samuel Kaski. Bayesian solutions to the label switching problem. *In Advances in Intelligent Data Analysis VIII*, Lecture Notes in Computer Science, pages 381–392, Berlin / Heidelberg, 2009. Springer.
30. Rastas, Pasi. A general framework for local pairwise alignment statistics with gaps. *Algorithms in bioinformatics*. pp. 233-245.
31. Rizo, David; Inesta, Jose M; Lemström, Kjell. Ensemble of state-of-the-art methods for polyphonic music comparison. *Proceedings of the Workshop on Exploring Musical Information Spaces*, ECDL 2009, Corfu, Greece, October 2009. - cop. 2009. pp. 46-51.
32. Rizo, David; Lemström, Kjell; Inesta, Jose M. Tree representation in combined polyphonic music comparison. *Computer music modeling and retrieval*. pp. 177-195.
33. Rousu, Juho. Bioinformatics: Proceedings of AMICT'2008: Vol. 10 / editors: Jussi Kangasharju, Yury A. Bogoyavlenskiy. pp. 50-57.
34. Sirén, Jouni. Compressed suffix arrays for massive data. *String processing and information retrieval*. pp. 63-74.
35. Eerika Savia, Kai Puolamäki, and Samuel Kaski. Two-way grouping by one-way topic models. *In Advances in Intelligent Data Analysis VIII*, Lecture Notes in Computer Science, pages 178–189, Berlin / Heidelberg, 2009. Springer.
36. S. Shimizu, A. Hyvärinen, Y. Kawahara and T. Washio. A direct method for estimating a causal ordering in a linear non-Gaussian acyclic model. *UAI 2009*.
37. Mika Sulkava. Environmental informatics and data mining in analysis of CO2 data. *In Novel Data Mining Strategies for Exploring Biogeochemical Cycles and Biosphere-Atmosphere Interactions – Workshop*, pages 19–20, Jena, Germany, June 2009.
38. Janne Toivola and Jaakko Hollmén. Feature extraction and selection from vibration measurements for structural health monitoring. *In Advances in Intelligent Data Analysis VIII*, volume 5772 of LNCS, pages 213–224, Berlin / Heidelberg, August 2009. Springer.
39. Antti Ukkonen. *Encyclopedia of Database Systems*, chapter Frequent Partial Orders. Springer, 2009.
40. Antti Ukkonen. Mining local correlation patterns in sets of sequences. *In Discovery Science, 12th International Conference, DS 2009*, pages 347–361, 2009.
41. Roman Yangarber, Ralf Steinberger. Automatic Epidemiological Surveillance from On-line News in MedISys and PULS. In Proceedings of IMED-2009: International Meeting on Emerging Diseases and Surveillance (2009) Vienna, Austria.
42. Yu, Huizhen; Bertsekas, Dimitri P. Basis function adaptation methods for cost approximation in MDP. *IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. - IEEE, 2009. 8 pp.
43. K. Zhang and A. Hyvärinen. On the Identifiability of the Post-Nonlinear Causal Model. *UAI 2009*.
44. K. Zhang and A. Hyvärinen. Causality Discovery with Additive Disturbances: An Information-Theoretical Perspective. *ECML 2009*.

45. Zhou, Fang; Mahler, Sebastien; Toivonen, Hannu. Review of network abstraction techniques. *Proceedings of the Workshop on Explorative Analytics of Information Networks 2009*. 2009. pp. 50-63.
46. Åstrand, Matti; Floréen, Patrik; Polishchuk, Valentin; Rybicki, Joel; Suomela, Jukka; Uitto, Jara. A local 2-approximation algorithm for the vertex cover problem. *Distributed computing*. pp. 191-205.

### Technical reports and other publications

1. Antti Ajanki, Mark Billingham, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, and Timo Tossavainen. Ubiquitous contextual information access with proactive retrieval and augmentation. Technical Report TKK-ICS-R27, *Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, December 2009*.
2. Astikainen, Katja; Holm, Liisa; Pitkänen, Esa; Rousu, Juho. Reaction kernels. *University of Helsinki, Department of Computer Science, 2009*. - 13 pp.
3. Bertsekas, Dimitri P; Yu, Huizhen. A unifying polyhedral approximation framework for convex optimization /. - USA: *Massachusetts Institute of Technology, 2009*. - 30 pp.
4. De Raedt, Luc; Kimmig, Angelika; Gutmann, Bernd; Kersting, Kristian; Santos Costa, Vitor; Toivonen, Hannu. Probabilistic inductive querying using ProbLog. *Katholieke Universiteit Leuven, Department of Computer Science 2009*, 29 pp.
5. Dzeroski, Saso; Geurts, Pierre; Rousu, Juho. Machine learning in systems biology: Helsinki University Printing House, 2009. - 180 pp.
6. Sami Hanhijärvi, Kai Puolamäki, and Gemma C. Garriga. Multiple hypothesis testing in pattern discovery. Technical Report TKK-ICS-R21, *Helsinki University of Technology, Department of Information and Computer Science, November 2009*.
7. Paula Järvinen, Kai Puolamäki, Pekka Siltanen, and Markus Ylikerälä. Visual analytics. Final report. VTT Working Papers 117, *VTT Technical Research Centre of Finland, Espoo, Finland, 2009*.
8. Kucherov, Gregory; Ukkonen, Esko, editors. Combinatorial pattern matching :. - Berlin : Springer, cop. 2009. - xiii, 370 pp.
9. Kai Puolamäki, editor. Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery, Otaniemi, 2009.

## 2008

### Articles in refereed scientific journals

1. Alava, Mikko; Ardelius, John; Aurell, Erik; Kaski, Petteri; Krishnamurthy, Supriya; Orponen, Pekka; Seitz, Sakari. Circumspect descent prevails in solving random constraint satisfaction problems. *Proceedings of the National Academy of Sciences of the United States of America*. 105 (2008): 40, pp. 15253-15257.
2. Astikainen, Katja; Holm, Liisa; Pitkänen, Esa; Szedmak, Sándor; Rousu, Juho. Towards structured output prediction of enzyme function. *BMC proceedings*. 2(Suppl 4) (2008): S2, 10 pp.
3. Böcker, Sebastian; Mäkinen, Veli. Combinatorial approaches for mass spectra recalibration. *IEEE/ACM transactions on computational biology and bioinformatics*. 5 (2008): 1, pp. 91-100.
4. Chee, Yeow Meng; Kaski, Petteri. An enumeration of graphical designs. *Journal of combinatorial designs*. 16 (2008), pp. 70-85.

5. De Raedt, Luc; Kersting, Kristian; Kimmig, Angelica; Revoredo, Kate; Toivonen, Hannu. Compressing probabilistic Prolog programs. *Machine learning*. 70 (2008): 2-3, pp. 151-168.
6. Dementiev, Roman; Kärkkäinen, Juha; Mehnert, Jens; Sanders, Peter. Better external memory suffix array construction. *Journal of experimental algorithmics*, JEA. 12 (2008): 3.4, pp. 3.4:1-3.4:24.
7. Garriga, Gemma C.; Kralj, Petra; Lavrac, Nada. Closed Sets for Labeled Data. *J. Mach. Learn. Res.*, 2008. Vol. 9, pp. 559-580.
8. Giegerich, Robert; Brazma, Alvis; Jonassen, Inge; Ukkonen, Esko; Vingron, Martin. The BREW workshop series. *Briefings in bioinformatics*. - 9 (2008): 3, pp. 250-253.
9. Haiminen, Niina; Gionis, Aristides; Laasonen, Kari. Algorithms for unimodal segmentation with applications to unimodality detection. *Knowledge and information systems*. b 14 (2008): 1, pp. 39-57.
10. Haiminen, Niina; Mannila, Heikki; Terzi, Evimaria. Determining significance of pairwise co-occurrences of events in bursty sequences. *BMC bioinformatics*. 9 (2008): 336, 10 pp.
11. Heinonen, Markus; Rantanen, Ari; Mielikainen, Taneli; Kokkonen, Juha; Kiuru, Jari; Ketola, Raimo A.; Rousu, Juho. FiD. *Rapid communications in mass spectrometry*. 22 (2008): 19, pp. 3043-3052.
12. Henriksson, Linda; Nurminen, Lauri; Hyvärinen, Aapo; Vanni, Simo. Spatial frequency tuning in human retinotopic visual areas. *Journal of vision*. 8 (2008) :10, pp. 1-13.
13. Hintsanen, Petteri; Toivonen, Hannu. Finding reliable subgraphs from large probabilistic graphs. *Data mining and knowledge discovery*. 17 (2008): 1, pp. 3-23.
14. Hoyer, Patrik O; Shimizu, Shohei; Kerminen, Antti; Palviainen, Markus. Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *International journal of approximate reasoning*. 49 (2008): 2, pp. 362-378.
15. Hyvärinen, Aapo. Optimal approximation of signal priors. *Neural computation*. 20 (2008): 12, pp. 3087-3110.
16. Kaban, Ata; Bingham, Ella. Factorisation and denoising of 0-1 data. *Neurocomputing*. 71 (2008): 10-12, pp. 2291-2308.
17. Kaski, Petteri; Östergård, Patric R. J. There are exactly five biplanes with  $k = 11$ . *Journal of combinatorial designs*. 16 (2008), pp. 117-127.
18. Kaski, Petteri; Östergård, Patric R. J; Topalova, Svetlana; Zlatarski, Rosen. Steiner triple systems of order 19 and 21 with subsystems of order 7. *Discrete mathematics*. 308 (2008): 13, pp. 2732-2741.
19. Lindgren, Jussi T; Hurri, Jarmo; Hyvärinen, Aapo. Spatial dependencies between local luminance and contrast in natural images. *Journal of vision*. 8 (2008) :12, pp. 1-13.
20. Liow, Lee Hsiang; Fortelius, Mikael; Bingham, Ella; Lintulaakso, Kari; Mannila, Heikki; Flynn, Larry; Stenseth, Nils Chr. Higher origination and extinction rates in larger mammals. *Proceedings of the National Academy of Sciences of the United States of America*. 105 (2008): 16, pp. 6097-6102.
21. Liow, Lee Hsiang; Fortelius, Mikael; Bingham, Ella; Lintulaakso, Kari; Mannila, Heikki; Flynn, Larry; Stenseth, Nils Chr. Reply to Vilar et al. *Proceedings of the National Academy of Sciences of the United States of America*. 105 (2008): 35, pp. E57.
22. Lundmark, P.E.; Liljedahl, U.; Boomsma, D.I.; Mannila, Heikki; Martin, N.G.; Palotie, A.; Peltonen, L.; Perola, M.; Spector, T.D.; Syvanen, A.-C. Evaluation of HapMap data in six populations of European descent. *European Journal of Human Genetics*, 2008. pp. 1-9.
23. Miettinen, Pauli. On the Positive-Negative Partial Set Cover problem. *Information processing letters*. 108 (2008): 4, pp. 219-221.
24. Miettinen, Pauli. The Boolean column and column-row matrix decompositions. *Data mining and knowledge discovery*. 17 (2008): 1-2, pp. 39-56.

25. Miettinen, Pauli; Mielikainen, Taneli; Gionis, Aristides; Das, Gautam; Mannila, Heikki. The discrete basis problem. *IEEE transactions on knowledge and data engineering*. 20 (2008): 10, pp. 1348-1362.
26. Mitchell, Joseph S. B; Polishchuk, Valentin. Minimum-perimeter enclosures. *Information processing letters*. 107 (2008), pp. 120-124.
27. Myllykangas, Samuel; Junnila, Siina; Kokkola, Arto; Autio, Reija; Scheinin, Ilari; Kiviluoto, Tuula; Karjalainen-Lindsberg, Marja-Liisa; Hollmén, Jaakko; Knuutila, Sakari; Puolakkainen, Pauli; Monni, Outi. Integrated Gene Copy Number and Expression Microarray Analysis of Gastric Cancer Highlights Potential Target Genes. *International Journal of Cancer*, 2008. Vol. 123, nro 4, pp. 817-825.
28. Myllykangas, Samuel; Tikka, Jarkko; Böhling, Tom; Knuutila, Sakari; Hollmén, Jaakko; Classification of human cancers based on DNA copy number amplification modeling. *BMC Medical Genomics*, 2008. Vol. 1, nro 15.
29. Mäkinen, Veli; Navarro, Gonzalo. Dynamic entropy-compressed sequences and full-text indexes. *ACM transactions on algorithms*. 4 (2008): 3, Art. 32.
30. Pitkänen, Esa; Akerlund, Arto; Rantanen, Ari; Jouhten, Paula; Ukkonen, Esko. ReMatch. *Journal of Integrative Bioinformatics*. 5 (2008): 102, 13 pp.
31. Puolamäki, Kai; Hanhijärvi, Sami; Garriga, Gemma C.; An approximation ratio for biclustering. *Information Processing Letters*, 2008. Vol. 108, nro 2, pp. 45-49.
32. Rantanen, Ari; Rousu, Juho; Jouhten, Paula; Zamboni, Nicola; Maaheimo, Hannu; Ukkonen, Esko. An analytic and systematic framework for estimating metabolic flux ratios from <sup>13</sup>C tracer experiments. *BMC bioinformatics*. 9 (2008): 266, 19 pp.
33. Ruosaari, Salla; Hienonen-Kempas, Tuija; Puustinen, Anne; Sarhadi, Virinder K.; Hollmén, Jaakko; Knuutila, Sakari; Saharinen, Juha; Wikman, Harriet; Anttila, Sisko. Pathways affected by asbestos exposure in normal and tumour tissue of lung cancer patients. *BMC Medical Genomics*, 2008. Vol. 1, nro 55,
34. Ruosaari, Salla; Nymark, Penny; Aavikko, Mervi; Kettunen, Eeva; Knuutila, Sakari; Hollmén, Jaakko; Norppa, Hannu; Anttila, Sisko; Aberrations of chromosome 19 in asbestos-associated lung cancer and in asbestos-induced micronuclei of bronchial epithelial cells in vitro. *Carcinogenesis*, 2008. Vol. 29, nro 5, pp. 913-917
35. Sevon, Petteri; Eronen, Lauri. Subgraph queries by context-free grammars. *Journal of Integrative Bioinformatics*. 5 (2008): 100, 16 pp.
36. Tatti, Nikolaj. Maximum entropy based significance of itemsets. *Knowledge and Information Systems*, 2008. Vol. 17, nro 1,
37. Tikka, Jarkko; Hollmén, Jaakko. A Sequential Input Selection Algorithm for Long-term prediction of Time Series. *Neurocomputing*, 2008. Vol. 71, nro 13-15, pp. 2604-2615.
38. Varjosalo, Markku; Björklund, Mikael.; Cheng, Fang; Syvanen, Heidi; Kivioja, Teemu; Kilpinen, Sami; Sun, Zairen; Kallioniemi, Olli; Stunnenberg, Hendrik G; He, Wei-Wu; Ojala, Päivi; Taipale, Jussi. Application of active and kinase-deficient kinome collection for identification of kinases regulating hedgehog signaling. *Cell*. 133 (2008): 3, pp. 537-548.
39. Wrage, Michaela; Ruosaari, Salla; Eijk, Paul P.; Kaifi, Jussuf T.; Hollmén, Jaakko; Yekebas, Emre F.; Izbicki, Jacob R.; Brakenhoff, Ruud H.; Streichert, Thomas; Riethdorf, Sabine; Ylstra, Bauke; Pantel, Klaus; Wikman, Harriet. Genomic Profiles Associated with Early Micrometastasis in Lung Cancer: Relevance of 4q Deletion. *Clinical Cancer Research*, 2008.

40. Yu, Huizhen; Bertsekas, Dimitri P. On near optimality of the set of finite-state controllers for average cost POMDP. *Mathematics of operations research*. 33 (2008): 1, 1-11 pp.

*Refereed conference articles and articles in edited books*

1. Ahonen, Teppo E; Lemström, Kjell. Identifying cover songs using normalized compression distance. *International Workshop on Machine Learning and Music, MML'08*, July 5-12, 2008, Helsinki, Finland, 2 pp.
2. Ala-Härkönen, Väinö; Brunberg, Jussi; Lemström, Kjell; Mikkilä, Niko. JMIR serves Mozart. *CMMR 2008*, pp. 148-157.
3. Arkin, Esther M; Mitchell, Joseph S. B; Polishchuk, Valentin. Maximum thick paths in static and dynamic environments. *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry (SCG-08)*, June 9-11, 2008, College Park, Maryland, USA, pp. 20-27.
4. Björklund, Andreas; Husfeldt, Thore; Kaski, Petteri; Koivisto, Mikko. Computing the Tutte polynomial in vertex-exponential time. *49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 677-686.
5. Björklund, Andreas; Husfeldt, Thore; Kaski, Petteri; Koivisto, Mikko. The travelling salesman problem in bounded degree graphs. *Automata, languages and programming*, pp. 198-209.
6. Doucet, Antoine; Lehtonen, Miro. Let's phrase it. *Proceedings of the SIGIR 2008 Workshop on Focused Retrieval*, 24 July, 2008, Singapore, pp. 9-14.
7. Efrat, Alon; Fekete, Sándor P; Gaddehosur, Poornananda R; Mitchell, Joseph S. B; Polishchuk, Valentin; Suomela, Jukka. Improved approximation algorithms for relay placement. *Algorithms - ESA 2008*, pp. 356-367.
8. Fischer, Johannes; Mäkinen, Veli; Navarro, Gonzalo. An(other) entropy-bounded compressed suffix tree. *Combinatorial pattern matching*, pp. 152-165.
9. Fischer, Johannes; Mäkinen, Veli; Välimäki, Niko. Space-efficient string mining under frequency constraints. *Eighth IEEE International Conference on Data Mining, ICDM 2008*, 15-19 December, 2008, Pisa, Italy: IEEE Computer Society 2008, pp.193-202.
10. Floréen, Patrik; Hassinen, Marja; Kaski, Petteri; Suomela, Jukka. Tight local approximation results for max-min linear programs. *Algorithmic aspects of wireless sensor networks*, pp. 2-17.
11. Floréen, Patrik; Kaski, Petteri; Musto, Topi; Suomela, Jukka. Approximating max-min linear programs with local algorithms. *proceedings of the 2008 IEEE International Parallel and Distributed Processing Symposium*, April 14-18, 2008, Miami, Florida: IEEE 2008, 10 pp.
12. Floréen, Patrik; Kaski, Petteri; Musto, Topi; Suomela, Jukka. Local approximation algorithms for scheduling problems in sensor networks. *Algorithmic aspects of wireless sensor networks*, pp. 99-113.
13. Gallo, Arianna; Miettinen, Pauli; Mannila, Heikki. Finding subgroups having several descriptions. *Proceedings of the Eighth SIAM International Conference on Data Mining*, April 24-26, 2008, Atlanta, Georgia, pp. 334-345.
14. Garriga, Gemma C.; Ukkonen, Antti; Mannila, Heikki. Feature Selection in Taxonomies with Applications to Paleontology. *Discovery Science, 11th International Conference, DS 2008*, Budapest, Hungary, October 13-16, 2008. pp. 112-123.
15. Garriga, Gemma C; Junttila, Esa; Mannila, Heikki. Banded structure in binary matrices. *KDD 08*, pp. 292-300.

16. Goethals, Bart; Page, Wim Le; Mannila, Heikki; Mining Association Rules of Simple Conjunctive Queries. *SIAM Data Mining Conference 2008*. pp. 96-107.
17. Grandoni, Fabrizio; Gupta, Anupam; Leonardi, Stefano; Miettinen, Pauli; Sankowski, Piotr; Singh, Mohit. Set covering with our eyes closed. *49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 347-356.
18. Hanhijärvi, Sami; Garriga, Gemma C.; Puolamäki, Kai. Randomization Techniques for Statistical Significance Testing on Graphs. *MLG-2008: 6th International Workshop on Mining and Learning with Graphs*, Helsinki, Finland, July 4-5, 2008.
19. Hassinen, Marja; Polishchuk, Valentin; Suomela, Jukka. Local 3-approximation algorithms for weighted dominating set and vertex cover in quasi unit-disk graphs. *Adjunct workshop proceedings: 2008*, 4 pp.
20. Hintsanen, Petteri; Toivonen, Hannu. Finding reliable subgraphs from large probabilistic graphs. *Machine learning and knowledge discovery in databases*, 15 pp.
21. Hollmén, Jaakko; Raiko, Tapani. Learning mixture models - courseware for finite mixture distributions of multivariate Bernoulli distributions. *In Proceedings of Teaching Machine Learning - workshop on open problems and new directions*. Saint-Etienne 2008,
22. Hoyer, Patrik O; Hyvärinen, Aapo; Scheines, Richard; Spirtes, Peter; Ramsey, Joseph; Lacerda, Gustavo; Shimizu, Shohei. Causal discovery of linear acyclic models with arbitrary distributions. *Uncertainty in Artificial Intelligence*, 8 pp.
23. Hyvärinen, Aapo. Estimation theory and information geometry based on denoising. *Proceedings of the First Workshop on Information Theoretic Methods in Science and Engineering*, August 18-20, 2008, Tampere, Finland, 8 pp.
24. Hyvärinen, Aapo; Shimizu, Shohei; Hoyer, Patrik O. Causal modelling combining instantaneous and lagged effects. *Proceedings of the 25th International Conference on Machine Learning*, 8 pp.
25. Hyvönen, Saara; Miettinen, Pauli; Terzi, Evimaria. Interpretable nonnegative matrix decompositions. *KDD 08*, pp. 345-353.
26. Karvonen, Mikko; Lemström, Kjell. Using mathematical morphology for geometric music information retrieval. *International Workshop on Machine Learning and Music*, MML'08, July 5-12, 2008, Helsinki, Finland. -, 2 pp.
27. Kim, Joondong; Mitchell, Joseph S. B; Polishchuk, Valentin; Vihavainen, Arto. Routing a maximum number of disks through a scene of moving obstacles. *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry (SCG-08)*, June 9-11, 2008, College Park, Maryland, USA, pp. 230-231.
28. Korpela, Mikko; Mäkinen, Harri; Sulkava, Mika; Nöjd, Pekka; Hollmén, Jaakko; Smoothed Prediction of the Onset of Tree Stem Radius Increase Based on Temperature Patterns. *The 11th International Conference on Discovery Science (DS-2008)*. 2008, Springer-Verlag, pp. 100-110.
29. Krasnoshchekov, Dmitry; Polishchuk, Valentin. Robust curve reconstruction with k-order  $\gamma$ -shapes. *Proceedings of the IEEE International Conference on Shape Modeling and Applications*, 4-6 June, 2008, Stony Brook, New York, USA: IEEE 2008, pp. 279-280.
30. Kärkkäinen, Juha. Suffix array construction. *Encyclopedia of Algorithms*, pp. 919-921.
31. Kärkkäinen, Juha; Rantala, Tommi. Engineering radix sort for strings\*. *String processing and information retrieval*, pp. 3-14.
32. Kärkkäinen, Juha; Ukkonen, Esko. Multidimensional string matching. *Encyclopedia of Algorithms*, pp. 559-562.

33. Lacerda, Gustavo; Spirtes, Peter; Ramsey, Joseph; Hoyer, Patrik O. Discovering cyclic causal models by Independent Components Analysis. *Uncertainty in Artificial Intelligence*, 9 pp.
34. Lehtonen, Miro; Doucet, Antoine. Phrase detection in the Wikipedia. *Focused access to XML documents*, pp. 115-121.
35. Lehtonen, Miro; Doucet, Antoine. XML-aided phrase indexing for hypertext documents. *SIGIR 2008*, pp. 843-844.
36. Lemström, Kjell; Mikkilä, Niko; Mäkinen, Veli. Fast index based filters for music retrieval. *ISMIR 2008*, pp. 677-682.
37. Lindgren, Jussi T; Hurri, Jarmo; Hyvärinen, Aapo. Unsupervised learning of dependencies between local luminance and contrast in natural images. *Proceedings of the International Joint Conference on Neural Networks, IJCNN2008*, June 1-6, 2008, Hong Kong, China: IEEE Xplore cop. 2008, pp. 356-362.
38. Lindgren, Jussi T; Hyvärinen, Aapo. On the learning of nonlinear visual features from natural images by optimizing response energies. *Proceedings of the International Joint Conference on Neural Networks, IJCNN2008*, June 1-6, 2008, Hong Kong, China: IEEE Xplore cop. 2008. , pp. 1026-1033.
39. Mahlberg, Riitta; Mannila, Juha; Romu, Jyrki; Nikkola, Juha; Ilola, Risto; Söderberg, Outi; Koskinen, Jari; Hannula, Simo-Pekka; Mahiout, Amar. Soil resistant and self-cleaning surfaces of stainless steel with new Sol-gel and ALD coatings. *The 6th European Stainless Steel Conference Science and Market 2008*, June 10-13, 2008, Helsinki, Finland. Jernkontoret 2008, pp. 101-108.
40. Mannila, Heikki. Finding Total and Partial Orders from Data for Seriation. *Discovery Science*. 2008, Springer, pp. 16-25.
41. Mavroeidis, Dimitrios; Bingham, Ella. Enhancing the stability of spectral ordering with sparsification and partial supervision. *Eighth IEEE International Conference on Data Mining, ICDM 2008*, 15-19, December 2008, Pisa, Italy: IEEE Computer Society 2008, 10 pp.
42. Merilä, Päivi; Derome, John; Luysaert, Sebastiaan; Sulkava, Mika; Hollmén, Jaakko; Mustajärvi, Kaisa; Nöjd, Pekka. How are N and S in deposition, in percolation water and in upper soil layers reflected in the chemical composition of needles in Finland?. *The Scientific Seminar on Forest Condition Monitoring and Ecosystem Functioning in Northern Europe under the Forest Focus and ICP Forests programmes*. Vantaa 2008, Finnish Forest Research Institute, pp. 39-42.
43. Miah, M.; Hristidis, V.; Das, Gautam; Mannila, Heikki. Standing Out in a Crowd: Selecting Attributes for Maximum Visibility. *International Conference on Data Engineering (ICDE 2008)*. pp. 356-365.
44. Mäkinen, Veli. Compressed suffix array. *Encyclopedia of Algorithms*, pp. 174-176.
45. Mäkinen, Veli; Navarro, Gonzalo. Compressed text indexing. *Encyclopedia of Algorithms*, pp. 176-178.
46. Mäkinen, Veli; Navarro, Gonzalo. On self-indexing images. *DCC 2008*, pp. 422-431.
47. Mäkinen, Veli; Ukkonen, Esko. Point pattern matching. *Encyclopedia of Algorithms*, pp. 657-660.
48. Nock, Richard; Luosto, Panu; Kivinen, Jyrki. Mixed Bregman clustering with approximation guarantees. *Machine learning and knowledge discovery in databases*, pp. 154-169.
49. Ojala, Markus; Vuokko, Niko; Kallio, Aleks; Haiminen, Niina; Mannila, Heikki. Randomization of real-valued matrices for assessing the significance of data mining results. *Proceedings of the Eighth SIAM International Conference on Data Mining*, April 24-26, 2008, Atlanta, Georgia, pp. 494-505.

50. Pitkänen, Esa; Rantanen, Ari; Rousu, Juho; Ukkonen, Esko. A computational method for reconstructing gapless metabolic networks. *Bioinformatics Research and Development*, pp. 288-302.
51. Polishchuk, Valentin; Suomela, Jukka. Optimal backlog in the plane. *Algorithmic aspects of wireless sensor networks*, pp. 141-150.
52. Puolamäki, Kai; Ajanki, Antti; Kaski, Samuel; Learning to learn implicit queries from gaze patterns. *ICML 2008, Twenty-Fifth International Conference on Machine Learning*. Madison 2008, pp. 760-767.
53. Raiko, Tapani; Puolamäki, Kai; Karhunen, Juha; Hollmén, Jaakko; Honkela, Antti; Kaski, Samuel; Mannila, Heikki; Oja, Erkki; Simula, Olli. Macadamia: Master's Programme in Machine Learning and Data Mining. *In Proceedings of Teaching Machine Learning - Workshop on Open Problems and New Directions*. Saint-'Etienne 2008.
54. Rastas, Pasi; Koivisto, Mikko; Mannila, Heikki; Ukkonen, Esko. Phasing genotypes using a hidden Markov model. *Bioinformatics algorithms*, pp. 355-372.
55. Rastas, Pasi; Kollin, Jussi; Koivisto, Mikko. Fast Bayesian haplotype inference via Context Tree Weighting. *Proceedings of the 8th International Workshop, WABI 2008*, Karlsruhe, Germany, September 15-19, 2008, pp. 259-270.
56. Rizo, David; Lemström, Kjell; Inesta, Jose M. Tree structured and combined methods for comparing metered polyphonic music. *CMMR 2008*, pp. 263-278.
57. Sirén, Jouni; Välimäki, Niko; Mäkinen, Veli; Navarro, Gonzalo. Run-length compressed indexes are superior for highly repetitive sequence collections. *String processing and information retrieval*, pp. 164-175.
58. Sirvio, Konsta; Hollmén, Jaakko. Spatio-Temporal Road Condition Forecasting With Markov Chains and Artificial Neural Networks. *Third International Workshop in Hybrid Artificial Intelligent Systems (HAIS'08)*. 2008, Springer-Verlag, pp. 204-211.
59. Steinberger, Ralf Flavio Fuart, Erik van der Groot, Clive Best, Peter von Etter, Roman Yangarber. Text Mining from the Web for Medical Intelligence. *In Mining Massive Data Sets for Security*, OIS Press. (2008)
60. Sulkava, Mika; Mäkinen, Harri; Nöjd, Pekka; Hollmén, Jaakko. Automatic detection of onset and cessation of tree stem radius increase using dendrometer data and CUSUM charts. *European Symposium on Time Series Prediction - ESTSP'08*. Porvoo 2008, Multiprint Oy / Otamedia, pp. 77-86.
61. Tatti, Nikolaj; Heikinheimo, Hannes. Decomposable Families of Itemsets. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2008)*, Antwerp, Belgium, Sept. 15-19, 2008. pp. 472-487.
62. Tatti, Nikolaj; Vreeken, Jilles. Finding Good Itemsets by Packing Data. *Eight IEEE International Conference on Data Mining (ICDM 2008)*.
63. Tikka, Jarkko; Hollmén, Jaakko. Selection of important input variables for RBF networks using partial derivatives. *The 16th European Symposium on Neural Networks (ESANN 2008)*. pp. 167-172.
64. Ukkonen, Antti; Castillo, Carlos; Donato, Debora; Gionis, Aristides. Searching the Wikipedia with Contextual Information. *Conference on Information and Knowledge Management, Proceedings of the 17th ACM conference on Information and knowledge mining*. 2008, ACM, New York, NY, USA, pp. 1351-1352.
65. Wong, Davin; Bingham, Ella; Hyvönen, Saara. Recommendation of multimedia items by link analysis and collaborative filtering. *Proceedings of the Second International Conference on Weblogs and Social Media*, March 30 - April 2, 2008, Seattle, Washington, USA: AAAI Press 2008, 2 pp.



66. Vuorio, Tiina; Jaakkola, Jyrki; Kolari, Mika; Mannila, Juha; Nikkola, Juha; Liu, Xuwen; Söderberg, Outi; Mahiout, Amar; Hannula, Simo-Pekka. Wear and chemical resistance of sol-gel coatings on the stainless steel substrate. *The 6th European Stainless Steel Conference Science and Market 2008*, June 10-13, 2008, Helsinki, Finland. Jernkontoret 2008, 93-100.
67. Yangarber, Roman; Peter von Etter, Ralf Steinberger. Content Collection and Analysis in the Domain of Epidemiology. *In Proceedings of DrMED 2008: International Workshop on Describing Medical Web Resources at MIE 2008: the 21st International Congress of the European Federation for Medical Informatics (2008)* Göteborg, Sweden
68. Yu, Huitzen; Bertsekas, Dimitri P. New error bounds for approximations from projected linear equations. *European Workshop on Reinforcement Learning 2008*, 30 June - 4 July 2008, Villeneuve d'Ascq, France: 2008. -, pp. 253-267.

#### *Technical reports and other publications*

1. Blohm, Sebastian, Ulf Brefeld, Felix Jungermann, Roman Yangarber (eds.). High-Level Information Extraction. *Proceedings of Workshop held at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases: ECML/PKDD 2008*. Antwerp, Belgium
2. Kaski, Petteri; Puttonen, Olli. libexact user's guide: HIIT, 2008.
3. Mäkinen, Veli; Lindén, Greger; Kujala, Inka. From data to knowledge: University of Helsinki, 2008. - 144 pp, ill;
4. Mäkinen, Veli; Navarro, Gonzalo; Sirén, Jouni; Välimäki, Niko. Run-length compressed indexes for repetitive sequence collections. *University of Helsinki, Department of Computer Science*, 2008. - 39 pp.
5. Mäkinen, Veli; Sirén, Jouni; Välimäki, Niko. Storage and retrieval of individual genomes and other repetitive sequence collections. *University of Helsinki, Department of Computer Science*, 2008. - 18 pp.
6. Poibeau, Thierry, Horacio Saggion, Roman Yangarber (eds.). Multi-source, Multilingual Information Extraction and Summarization. *Proceedings of the Second MMIES Workshop, held at the 22nd International Conference on Computational Linguistics, COLING-2008*. Manchester, United Kingdom
7. Puolamäki, Kai; Kaski, Samuel. Bayesian Solutions to the Label Switching Problem. Espoo: *Teknillinen korkeakoulu*, 2008. iii, 8 p. (TKK reports in information and computer science 7; TKK-ICS-R7).
8. Tatti, Nikolaj; Heikinheimo, Hannes. Decomposable Families of Itemsets. Espoo: *Teknillinen korkeakoulu*, 2008. iv, 16 p. (TKK reports in information and computer science 1; TKK-ICS-R1).
9. Yangarber, Roman; Marko Salmenkivi, Marjaana Väilisalo. A Database of the Uralic Language Family for Etymological Research. Technical Report C-2008-38. *University of Helsinki, Department of Computer Science, Series of Publications C*.
10. Yu, Huizhen; Bertsekas, Dimitri P. New error bounds for approximations from projected linear equations. *University of Helsinki, Department of Computer Science*, 2008. - 24 pp.

## PhD degrees and theses

### 2014

1. Galbrun, Esther. Methods for Redescription Mining. *University of Helsinki*.
2. Korhonen, Janne. Graph and Hypergraph Decompositions for Exact Algorithms. *University of Helsinki*.
3. Paalasmaa, Joonas. Monitoring Sleep with Force Sensor Measurement. *University of Helsinki*.

### 2013

1. Entner, Doris. Causal Structure Learning and Effect Identification in Linear Non-Gaussian Models and Beyond. *University of Helsinki*.
2. Eronen, Lauri. Computational Methods for Augmenting Associations-based Gene Mapping. *University of Helsinki*.
3. Hyttinen, Antti. Discovering Causal Relations in the Presence of Latent Confounders. *University of Helsinki*.
4. Lijffijt, Jeffrey. Computational methods for comparison and exploration of event sequences. *Aalto University*.
5. Luosto, Panu. Normalized Maximum Likelihood Methods for Clustering and Density Estimation. *University of Helsinki*.
6. Timonen, Mika. Term Weighting in Short Documents for Document Categorization, Keyword Extraction and Query Expansion. *University of Helsinki*.
7. Wettig, Johannes. Probabilistic, Information-Theoretic Models for Etymological Alignment. *University of Helsinki*.

### 2012

1. Hanhijärvi, Sami. Multiple hypothesis testing in data mining. *Aalto University*.
2. Heinonen, Markus. Computational Methods for Small Molecules. *University of Helsinki*.
3. Parviainen, Pekka. Algorithms for Exact Structure Discovery in Bayesian Networks. *University of Helsinki*.
4. Sirén, Jouni. Compressed Full-Text Indexes for Highly Repetitive Collections. *University of Helsinki*.
5. Vuokko, Niko. Testing the Significance of Patterns in Complex Null Hypotheses. *Aalto University*.
6. Välimäki, Niko. Applications of Compressed Data Structures on Sequences and Structured Data. *University of Helsinki*.
7. Wessman, Jaana. Mixture Model Clustering in the Analysis of Complex Diseases. *University of Helsinki*.
8. Zhou, Fang. Methods for Network Abstraction. *University of Helsinki*.

### 2011

1. Junttila, Esa. Patterns in Permuted Binary Matrices. *University of Helsinki*.
2. Hintsanen, Petteri. Simulation and Graph Mining Tools for Improving Gene Mapping Efficiency. *University of Helsinki*.
3. Ojala, Markus. Randomization Algorithms for Assessing the Significance of Data Mining Results. *Aalto University*.

### 2010

1. Heikinheimo, Hannes. Extending data mining techniques for frequent pattern discovery: trees, lowentropy sets, and crossmining. *Aalto University*.
2. Hämäläinen, Wilhelmiina. Efficient search for statistically significant dependency rules in binary data. *University of Helsinki*.

3. Kollin, Jussi. Computational Methods for Detecting Large-Scale Chromosome Rearrangements in SNP Data. *University of Helsinki*.
4. Lukk, Margus. Construction of a global map of human gene expression – the process, tools and analysis. *University of Helsinki*.
5. Pitkänen, Esa. Computational Methods for Reconstruction and Analysis of Genome-Scale Metabolic Networks. *University of Helsinki*.

## 2009

1. Hätönen, Kimmo. Data mining for telecommunications network log analysis. *University of Helsinki*
2. Kurki, Ilmari. Feature Integration in Human Vision. *University of Helsinki*
3. Köster, Urs. A Probabilistic approach to the primary visual cortex. *University of Helsinki*
4. Laasonen, Kari. Mining cell transition data. *University of Helsinki*
5. Makkonen, Juha. Semantic classes in topic detection and tracking. *University of Helsinki*
6. Miettinen, Pauli. Matrix decomposition methods for data mining. *University of Helsinki*
7. Rastas, Pasi. Computational Techniques for Haplotype Inference and for Local Alignment Significance. *University of Helsinki*

## 2008

1. Autio, Ilkka. Modeling efficient classification as a process of confidence assessment and delegation. *University of Helsinki*
2. Haiminen, Niina. Mining sequential data. *University of Helsinki*
3. Lindgren, Jussi T. Learning nonlinear visual processing from natural images. *University of Helsinki*
4. Ruosaari, Salla. Microarrays in lung cancer research: From comparative analyses to verified findings. *University of Helsinki*
5. Sulkava, Mika. Learning from environmental data: methods for analysis of forest nutrition time series. *Helsinki University of Technology*
6. Tikka, Jarkko. Input variable selection methods for construction of interpretable regression models. *Helsinki University of Technology*
7. Ukkonen, Antti. Algorithms for finding orders and analyzing sets of chains. *Helsinki University of Technology*

