# Probabilistic Automata: Semantics, Equivalence & Minimization

## Master Thesis

by

**Fabian Klopfer**

Universität Konstanz

Faculty of Sciences

Department of Computer and Information Science

**1st Reviewer:** Prof. Dr. Tatjana Petrov

**2nd Reviewer:** ToDo

Konstanz, 2020

**Abstract:**

Some New Abstract Text

# Contents

# 1 Introduction

# 2 Background

## 2.1 Algebraic Preliminaries

**Definition 2.1** (Equivalence Relation)**.**

**Definition 2.2** (Monoid)**.**

**Definition 2.3** (Ring, Semiring)**.**

**Theorem 2.3.1** (The Ring of Formal Power Series)**.**

## 2.2 Category Theory

## 2.3 Probability Theory

**Definition 2.4** (Sample space, events, $\sigma$-algebra, measurable space)**.** *Let $\Omega$ be a set, namely the set of possible outcomes of a chance experiment, called* **sample space***.*

*Let $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ with $\mathcal{P}$ the powerset, a set of subsets of the sample space, whose elements are called* **events***.*

*Let $\Omega$ be a sample space, $\mathcal{F}$ a set of events. $\mathcal{F}$ is called a* **$\sigma$-algebra** *over $\Omega$, if and only if*

- *the sample space is contained in the set of events,*

$$\Omega \in \mathcal{F}$$

- *the set of events is closed under complementation,*

$$A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$$

- *and the set of events is closed under countable union:*

$$\forall i \geq 0 : A_i \in F \Rightarrow ( \bigcup_{n \in \mathbb{N} A_n ) \in \mathcal{F}}$$

*The pair $(\Omega, \mathcal{F})$ is called a* **measurable space***.*

*Let $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ measurable spaces. A funtion $f : \Omega_1 \to \Omega_2$ is called a* **measurable function** *if and only if for every $A \in \mathcal{F}_2$ the preimage of $A$ under $f$ is in $\mathcal{F}_1$.*

$$\forall A \in \mathcal{F}_2 : f^{-1}(A) \in \mathcal{F}_1$$

**Definition 2.5** (Probability Space, Probability Measure, Discrete and Continuous Probability Space and Measure)**.**

**Definition 2.6** (Random Variable, Probability Distribution, Distribution Funtion and Cumulative Density Function)**.**

**Definition 2.7** (Mean, Median, Mode, Expectation, Variance)**.**

**Definition 2.8** (Joint Probability Distribution, Conditional Probability, Bayes Rule, Independence, Conditional Independence)**.**

**Definition 2.9** (Stochastic Process, Bernouilli & Binomial Process)**.**

**Definition 2.10** (Geometric Distribution)**.**

**Definition 2.11** ((Negative) Exponential Distr.)**.**

**Theorem 2.11.1** (Memoryless property)**.**

**Definition 2.12** (Markov Property, Markov Process, Time Homogeneity)**.**

## 2.4 Automata Theory

**Definition 2.13** (Transition System)**.**

**Definition 2.14** (Labelled Transition System)**.**

**Definition 2.15** (Path, Trace, Cylinder Sets, Prefix, Postfix)**.**

**Definition 2.16** (Determinism and Non-Determinism, Internal vs. External, Adversaries/Policies)**.**

**Definition 2.17** (Weighted Automata)**.**

**Definition 2.18** (Probabilistic Automata, Initial Distribution, transition probability function, stochastic matrix, transition probability matrix)**.**

**Remark** (Probabilistic vs. Non-Deterministic Choice)**.**

**Remark** (Sokolova's System Types Paper and focus on MC-like models)**.**

**Definition 2.19** (Markov Chains, Discrete-Time, Continuous-Time, Labelled, (MDP?, Generalized Stochastic Petri Nets?))**.**

**Definition 2.20** (Bisimulation Relations, Strong, weak, exact, ordinary, Prob., Buchholz)**.**

# 3 Semantics, Equivalence & Minimization

The general case, may help with some things, e.g. if proven independent of the semiring used. So for each subsection here starting with WAs independent of the Semiring that is used, continue with PA results that are as independent as possible from the concrete transition structure. Finally apply the aforementioned results to LMCs/MCs. Also discriminate between DTMC, MDP, PA model(s) I.e. per subsection apply the folling structure

**WA - General Case for arbitrary Semirings**

**PA - Results for other transition structures**

**MC-like models - Application of above and other literature on specific example of LMC**

## 3.1 Semantics

### 3.1.1 Parametrization and Initialization

### 3.1.2 Trace Semantics

Execution as Sparse Matrix Multiplication, (constrained) reachability (pr.), path/trace distributions, ergodicity, state residency time, Uniformization

### 3.1.3 Transient Semantics & Cylinder Sets of Executions (incl. Reward/weighted semantics)

wrt. probabilities?!

**Definition 3.1** (Transient probability distributiin)**.**

### 3.1.4 Threshold semantics

### 3.1.5 Funtion Transient Semantics (?)

word functions instead of probability functions?

## 3.2 Equivalence & Tanja's Draft Equivalences

### 3.2.1 Trace semantics is decidable in P for all

paz p.36, Kiefer , Tzeng, Schützenberger, Bollig & Zeitoun, Kiefer WA should hold for PA when using sufficient eps (theoretically unclean), Doyen,

### 3.2.2 Transient semantics is lower NP, upper EXP

As we need to construct the postfix space which is a cylinder set, transfers also to reward/weighted transient

### 3.2.3 Word Function-based

Same as above, just that lower bound is less; stop on first false; maybe eq of kiefer helps here

## 3.3 Minimization + wrt. Draft Eqs.

**Definition 3.2** (Lumping/lumpable).

### 3.3.1 Approaches

#### 3.3.1.1 Partition Refinement - Coalgebraic Approach

Proof: PA minimization is in P for uninitialized.
Deiffel, WiSSmann, Paz p.24ff (IntroToProbabilisticAutomata) for all in O(n log n) as partition refinement is minimal (see sources above) for uninitialized Automata.

#### 3.3.1.2 Schützenberger's Construction $\pm$ Arnoldi Iteration with Housholder Reflectors &

Show either that is also in P oder correct Kiefers runtime analysis. For uninitialized see kiefer; TODO figure out if kiefers reduction is flawed or if the runtime analysis of his algo is flawed.

### 3.3.2 Wrt. different semantics

#### 3.3.2.1 Decidability

#### 3.3.2.2 Complexity

# 4 Conclusions

# 5 References

# Bibliography


[1] *11.8. Statistics and execution plans - Chapter 11. Performance*. Jan. 3, 2020. URL: https://neo4j.com/docs/operations-manual/current/performance/statistics-execution-plans/ (visited on 01/11/2020).

[2] *6.4. Index Values and Order - Chapter 6. Query tuning*. Jan. 3, 2020. URL: https://neo4j.com/docs/cypher-manual/3.5/query-tuning/cypher-index-values-order/ (visited on 01/11/2020).

[3] *6.5. Planner hints and the USING keyword - Chapter 6. Query tuning*. Jan. 3, 2020. URL: https://neo4j.com/docs/cypher-manual/3.5/query-tuning/using/ (visited on 01/11/2020).

[4] *9th DIMACS Implementation Challenge: Shortest Paths*. June 14, 2010. URL: https://users.diag.uniroma1.it/challenge9/download.shtml (visited on 01/12/2020).

[5] Ryan P Adams. "Hierarchical Agglomerative Clustering". In: *Proc. Ninth SIAM Data Mining Conf.(SDM09)*. 2009, pp. 510–516.

[6] Renzo Angles. "The Property Graph Database Model." In: *AMW*. 2018.

[7] Mihael Ankerst et al. "OPTICS: Ordering Points To Identify the Clustering Structure". In: ACM Press, 1999, pp. 49–60.

[8] David Arthur and Sergei Vassilvitskii. "k-means++: The advantages of careful seeding". In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics. 2007, pp. 1027–1035.

[9] Geoffrey H Ball and David J Hall. "A clustering technique for summarizing multivariate data". In: *Behavioral science* 12.2 (1967), pp. 153–155.

[10] Michael Baron. *Probability and Statistics for Computer Scientists, Second Edition*. 2nd. Chapman & Hall/CRC, 2013.

[11] E. T. Bell. "Exponential Numbers". In: *The American Mathematical Monthly* 41.7 (1934), pp. 411–419. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2300300.

[12] Benjamin Bengfort, Rebecca Bilbro, and Tony Ojeda. *Applied Text Analysis with Python: Enabling Language-aware Data Products with Machine Learning*. " O'Reilly Media, Inc.", 2018.

[13] Pavel Berkhin. "A survey of clustering data mining techniques". In: *Grouping multidimensional data*. Springer, 2006, pp. 25–71.

[14] Tim Bock. *What is a Dendrogram? How to use Dendrograms | Displayr*. 2018. URL: https://www.displayr.com/what-is-dendrogram/ (visited on 11/26/2019).

[15] Annina Breen. *File:Taxonomic Rank Graph.svg - Wikimedia Commons*. 2015. URL: https://commons.wikimedia.org/wiki/File:Taxonomic_Rank_Graph.svg (visited on 11/23/2019).

[16] Jerome Bruner. *A study of thinking*. Routledge, 2017.

[17] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. "Density-based clustering based on hierarchical density estimates". In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer. 2013, pp. 160–172.

[18] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. "Density-based clustering based on hierarchical density estimates". In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer. 2013, pp. 160–172.

[19] Gunnar Carlsson. "Topology and data". In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 255–308.

[20] Claudio Carpineto and Giovanni Romano. "Mining Short-Rule Covers in Relational Databases". In: *Computational Intelligence* 19.3 (2003), pp. 215–234.

[21] Claudio Carpineto, Giovanni Romano, and Fondazione Ugo Bordoni. "Exploiting the potential of concept lattices for information retrieval with CREDO." In: *J. UCS* 10.8 (2004), pp. 985–1013.

[22] Tony F. Chan, Gene H. Golub, and Randall J. LeVeque. "Algorithms for Computing the Sample Variance: Analysis and Recommendations". In: *The American Statistician* 37.3 (1983), pp. 242–247. ISSN: 00031305. URL: http://www.jstor.org/stable/2683386.

[23] Kamalika Chaudhuri and Sanjoy Dasgupta. "Rates of convergence for the cluster tree". In: *NIPS*. 2010.

[24] Peter Cheeseman et al. "Autoclass: A Bayesian classification system". In: *Machine learning proceedings 1988*. Elsevier, 1988, pp. 54–64.

[25] James E Corter and Mark A Gluck. "Explaining basic categories: Feature predictability and information." In: *Psychological Bulletin* 111.2 (1992), p. 291.

[26] Mark Devaney and Ashwin Ram. "Efficient feature selection in conceptual clustering". In: *ICML*. Vol. 97. 1997, pp. 92–97.

[27] J. C. Dunn. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". In: *Journal of Cybernetics* 3.3 (1973), pp. 32–57. DOI: 10.1080/01969727308546046. eprint: https://doi.org/10.1080/01969727308546046. URL: https://doi.org/10.1080/01969727308546046.

[28] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: AAAI Press, 1996, pp. 226–231.

[29] Ludwig Fahrmeir et al. *Statistik: Der weg zur datenanalyse*. Springer-Verlag, 2016.

[30] Doug Fisher. "Iterative optimization and simplification of hierarchical clusterings". In: *Journal of artificial intelligence research* 4 (1996), pp. 147–178.

[31] Douglas Fisher and Pat Langley. *Approaches to Conceptual Clustering.* Tech. rep. CALIFORNIA UNIV IRVINE DEPT OF INFORMATION and COMPUTER SCIENCE, 1985.

[32] Douglas H. Fisher. "Knowledge acquisition via incremental conceptual clustering". In: *Machine Learning* 2.2 (Sept. 1987), pp. 139–172. ISSN: 1573-0565. DOI: 10.1007/BF00114265. URL: https://doi.org/10.1007/BF00114265.

[33] Jean-Gabriel Ganascia. "CHARADE: A Rule System Learning System." In: *IJCAI*. Vol. 87. 1987, pp. 234–239.

[34] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.

[35] John H Gennari, Pat Langley, and Doug Fisher. "Models of incremental concept formation". In: *Artificial intelligence* 40.1-3 (1989), pp. 11–61.

[36] Roger Germundsson and Eric W Weisstein. *XOR From MathWorld*. URL: http://mathworld.wolfram.com/XOR.html (visited on 12/03/2019).

[37] M Gluck. "Information, uncertainty and the utility of categories". In: *Proc. of the Seventh Annual Conf. on Cognitive Science Society, 1985*. 1985.

[38] Robert Godin, Hafedh Mili, et al. "Building and maintaining analysis-level class hierarchies using galois lattices". In: *OOPSLA*. Vol. 93. 1993, pp. 394–410.

[39] Robert Godin, Rokia Missaoui, and Hassan Alaoui. "INCREMENTAL CONCEPT FORMATION ALGORITHMS BASED ON GALOIS (CONCEPT) LATTICES". In: *Computational Intelligence* 11.2 (1995), pp. 246–267. DOI: 10.1111/j.1467-8640.1995.tb00031.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8640.1995.tb00031.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.1995.tb00031.x.

[40] Robert Godin, Rokia Missaoui, and Alain April. "Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods". In: *International Journal of Man-Machine Studies* 38.5 (1993), pp. 747–767.

[41] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[42] Keith Henderson et al. "It's who you know: graph mining using recursive structural features". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 663–671.

[43] Keith Henderson et al. "Rolx: structural role extraction & mining in large graphs". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2012, pp. 1231–1239.

[44] Matthew D Hoffman et al. "Stochastic variational inference". In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 1303–1347.

[45] Dmitry I Ignatov, Sergei O Kuznetsov, and Jonas Poelmans. "Concept-based biclustering for internet advertisement". In: *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE. 2012, pp. 123–130.

[46] Dmitry I. Ignatov. "Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields". In: *CoRR* abs/1703.02819 (2017). arXiv: 1703.02819. URL: http://arxiv.org/abs/1703.02819.

[47] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. "Data clustering: a review". In: *ACM computing surveys (CSUR)* 31.3 (1999), pp. 264–323.

[48] Xin Jin and Jiawei Han. "K-Means Clustering". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 563–564. ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_425. URL: https://doi.org/10.1007/978-0-387-30164-8_425.

[49] Kirthevasan Kandasamy et al. "Neural architecture search with bayesian optimisation and optimal transport". In: *Advances in Neural Information Processing Systems*. 2018, pp. 2016–2025.

[50] Teuvo Kohonen. "Exploration of very large databases by self-organizing maps". In: *Proceedings of International Conference on Neural Networks (ICNN'97)*. Vol. 1. IEEE. 1997, PL1–PL6.

[51] Teuvo Kohonen. "The self-organizing map". In: *Proceedings of the IEEE* 78.9 (1990), pp. 1464–1480.

[52] Sven Kosub. "A note on the triangle inequality for the Jaccard distance". In: *CoRR* abs/1612.02696 (2016). arXiv: 1612.02696. URL: http://arxiv.org/abs/1612.02696.

[53] Helmut Krcmar. "Informationsmanagement". In: *Informationsmanagement*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 85–111. ISBN: 978-3-662-45863-1. DOI: 10.1007/978-3-662-45863-1_4. URL: https://doi.org/10.1007/978-3-662-45863-1_4.

[54] Rudolf Kruse et al. *Computational intelligence: a methodological introduction*. Springer, 2016.

[55] Pat Langley et al. *An integrated cognitive architecture for autonomous agents*. Tech. rep. CALIFORNIA UNIV IRVINE SCHOOL OF INFORMATION and COMPUTER SCIENCE, 1990.

[56] Gang Leng, Thomas Martin McGinnity, and Girijesh Prasad. "Design for self-organizing fuzzy neural networks based on genetic algorithms". In: *IEEE Transactions on Fuzzy Systems* 14.6 (2006), pp. 755–766.

[57] Robert Levinson. "A self-organizing retrieval system for graphs." In: *AAAI*. 1984, pp. 203–206.

[58] Stuart Lloyd. "Least squares quantization in PCM". In: *IEEE transactions on information theory* 28.2 (1982), pp. 129–137.

[59] Christopher J MacLellan et al. "Trestle: A model of concept formation in structured domains". In: *Advances in Cognitive Systems* 4 (2016), pp. 131–150.

[60] James MacQueen et al. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.

[61] Leland McInnes and John Healy. "Accelerated hierarchical density based clustering". In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE. 2017, pp. 33–42.

[62] Kathleen McKusick and Kevin Thompson. "Cobweb/3: A portable implementation". In: (1990).

[63] Ryszard Michalski. "Knowledge acquisition through conceptual clustering: A theoretical framework and algorithm for partitioning data into conjunctive concepts". In: *International Journal of Policy Analysis and Information Systems* 4 (1980), pp. 219–243.

[64] Ryszard S Michalski and Robert E Stepp. "Learning from observation: Conceptual clustering". In: *Machine learning*. Springer, 1983, pp. 331–363.

[65] Tomas Mikolov et al. "Distributed Representations of Words and Phrases and their Compositionality". In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges et al. Curran Associates, Inc., 2013, pp. 3111–3119. URL: `http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf`.

[66] Boris Mirkin. *Mathematical classification and clustering*. Vol. 11. Springer Science & Business Media, 2013.

[67] Tom M Mitchell. "Generalization as search". In: *Artificial intelligence* 18.2 (1982), pp. 203–226.

[68] Tom Michael Mitchell. *The discipline of machine learning*. Vol. 9. Carnegie Mellon University, School of Computer Science, Machine Learning, 2006.

[69] Fionn Murtagh. "Ultrametric and generalized ultrametric in computational logic and in data analysis". In: *arXiv preprint arXiv:1008.3585* (2010).

[70] Fionn Murtagh and Pedro Contreras. "Hierarchical Clustering for Finding Symmetries and Other Patterns in Massive, High Dimensional Datasets". In: *Data Mining: Foundations and Intelligent Paradigms* (2012), pp. 95–130. ISSN: 1868-4408. DOI: `10.1007/978-3-642-23166-7_5`. URL: `http://dx.doi.org/10.1007/978-3-642-23166-7_5`.

[71] Thomas Neumann and Guido Moerkotte. "Characteristic sets: Accurate cardinality estimation for RDF queries with multiple joins". In: *2011 IEEE 27th International Conference on Data Engineering*. IEEE. 2011, pp. 984–994.

[72] Allen E. Nix and Michael D. Vose. “Modeling genetic algorithms with Markov chains”. In: *Annals of Mathematics and Artificial Intelligence* 5.1 (Mar. 1992), pp. 79–88. ISSN: 1573-7470. DOI: 10.1007/BF01530781. URL: https://doi.org/10.1007/BF01530781.

[73] Andrei Novikov. “PyClustering: Data Mining Library”. In: *Journal of Open Source Software* 4.36 (Apr. 2019), p. 1230. DOI: 10.21105/joss.01230. URL: https://doi.org/10.21105/joss.01230.

[74] G Oosthuizen and DR McGregor. “Induction through knowledge base normalisation”. In: *Proceedings of the 8th European Conference on Artificial Intelligence*. Pitman Publishing, Inc. 1988, pp. 396–401.

[75] Mateusz Pawlik and Nikolaus Augsten. “Tree edit distance: Robust and memory-efficient”. In: *Information Systems* 56 (2016), pp. 157–173.

[76] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[77] Jonas Poelmans et al. “Text mining scientific papers: a survey on FCA-based information retrieval research”. In: *Industrial Conference on Data Mining*. Springer. 2012, pp. 273–287.

[78] Beatriz Pontes, Raúl Giráldez, and Jesús S. Aguilar-Ruiz. “Biclustering on expression data: A review”. In: *Journal of Biomedical Informatics* 57 (2015), pp. 163–180. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2015.06.028. URL: http://www.sciencedirect.com/science/article/pii/S1532046415001380.

[79] Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. 3rd ed. New York, NY, USA: McGraw-Hill, Inc., 2003.

[80] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph databases*. " O'Reilly Media, Inc.", 2013.

[81] Eleanor Rosch et al. “Basic objects in natural categories”. In: *Cognitive Psychology* 8 (1976), pp. 382–439.

[82] SIGKDD. *SIGKDD*. Dec. 11, 2019. URL: https://www.kdd.org/ (visited on 12/11/2019).

[83] Herbert A Simon. *The sciences of the artificial*. 1968.

[84] P. H. A. Sneath. “The Application of Computers to Taxonomy”. In: *Microbiology* 17.1 (1957), pp. 201–226. ISSN: 1350-0872. DOI: https://doi.org/10.1099/00221287-17-1-201. URL: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-17-1-201.

[85] Dagobert Soergel. “Mathematical analysis of documentation systems: An attempt to a theory of classification and search request formulation”. In: *Information storage and retrieval* 3.3 (1967), pp. 129–173.

[86] Anselm Spoerri. “InfoCrystal: A visual tool for information retrieval & management”. In: *Proceedings of the second international conference on Information and knowledge management*. ACM. 1993, pp. 11–20.

[87] Kuo-Chung Tai. “The Tree-to-Tree Correction Problem”. In: *J. ACM* 26.3 (July 1979), pp. 422–433. ISSN: 0004-5411. DOI: 10.1145/322139.322143. URL: http://doi.acm.org/10.1145/322139.322143.

[88] Sergios Theodoridis and Konstantinos Koutroumbas. “Chapter 12 - Clustering Algorithms I: Sequential Algorithms”. In: *Pattern Recognition (Fourth Edition)*. Ed. by Sergios Theodoridis and Konstantinos Koutroumbas. Fourth Edition. Boston: Academic Press, 2009, pp. 627–652. ISBN: 978-1-59749-272-0. DOI: https://doi.org/10.1016/B978-1-59749-272-0.50014-1. URL: http://www.sciencedirect.com/science/article/pii/B9781597492720500141.

[89] Sergios Theodoridis and Konstantinos Koutroumbas. "Chapter 14 - Clustering Algorithms III: Schemes Based on Function Optimization". In: *Pattern Recognition (Fourth Edition)*. Ed. by Sergios Theodoridis and Konstantinos Koutroumbas. Fourth Edition. Boston: Academic Press, 2009, pp. 701–763. ISBN: 978-1-59749-272-0. DOI: https://doi.org/10.1016/B978-1-59749-272-0.50016-5. URL: http://www.sciencedirect.com/science/article/pii/B9781597492720500165.

[90] Kevin Thompson and Pat Langley. "Concept formation in structured domains". In: *Concept Formation*. Elsevier, 1991, pp. 127–161.

[91] Kevin Thompson and Pat Langley. "Incremental concept formation with composite objects". In: *Proceedings of the sixth international workshop on Machine learning*. Elsevier. 1989, pp. 371–374.

[92] Charles D. Tupper. "21 - Object and Object/Relational Databases". In: *Data Architecture*. Ed. by Charles D. Tupper. Boston: Morgan Kaufmann, 2011, pp. 369–383. ISBN: 978-0-12-385126-0. DOI: https://doi.org/10.1016/B978-0-12-385126-0.00021-8. URL: http://www.sciencedirect.com/science/article/pii/B9780123851260000218.

[93] Joe H Ward Jr. "Hierarchical grouping to optimize an objective function". In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244.

[94] Alfred North Whitehead and Bertrand Russell. *Principia mathematica*. Vol. 2. University Press, 1912.

[95] Rudolf Wille. "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts". In: *Ordered Sets*. Springer, 1982, pp. 445–470.

[96] David Wishart. "256. Note: An algorithm for hierarchical classifications". In: *Biometrics* (1969), pp. 165–170.

[97] Zhipeng Xie et al. "Concept lattice based composite classifiers for high predictability". In: *Journal of Experimental & Theoretical Artificial Intelligence* 14.2-3 (2002), pp. 143–156.

[98] Rui Xu and Donald C Wunsch. "Survey of clustering algorithms". In: (2005).

[99] Jason Yosinski et al. "Understanding Neural Networks Through Deep Visualization". In: *Deep Learning Workshop, International Conference on Machine Learning (ICML)*. 2015.

[100] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. "BIRCH: An Efficient Data Clustering Method for Very Large Databases". In: *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*. SIGMOD '96. Montreal, Quebec, Canada: ACM, 1996, pp. 103–114. ISBN: 0-89791-794-4. DOI: 10.1145/233269.233324. URL: http://doi.acm.org/10.1145/233269.233324.