

## Examples for Project Topics

- 1) Out-of-core data structures for priority queues, dictionaries, hashing, etc.
- 2) Out-of-core algorithms for the construction of suffix arrays and suffix trees
- 3) Out-of-core algorithms for graph problems
- 4) Out-of-core algorithms for clustering
- 5) Out-of-core algorithms for sequence analysis
- 6) Biological data compression (DNA and protein sequences, for example)
- 7) Biological data reduction techniques
- 8) Out-of-core algorithms for rules mining
- 9) Out-of-core algorithms for matrix problems
- 10) Out-of-core algorithms for motif search (planted motif search, for example)
- 11) Efficient algorithms for k-mer counting
- 12) Efficient algorithms for sequence assembly
- 13) Efficient algorithms for the closest pair problem
- 14) Xiaodong Yu, Kaixi Hou, Hao Wang, and Wu-chun Feng, Hierarchical Automata Construction for Approximate Pattern Matching on Automata Processors, IEEE Big Data 2017.
- 15) Ichitaro Yamazaki, Stanimire Tomov, and Jack Dongarra, Sampling Algorithms to Update Truncated SVD, IEEE Big Data 2017.
- 16) Tong Yang, Binchao Yin, Hang Li, Muhammad Shahzad, Steve Uhlig, Bin Cui, and Xiaoming Li, Rectangular Hash Table: Bloom Filter and Bitmap Assisted Hash Table with High Speed, IEEE Big Data 2017.
- 17) Yueyao Wang, Qinmin Vivian Hu, Yang Song, and Liang He, Potentiality of Healthcare Big data: Improving Search by Automatic Query Reformulation, IEEE Big Data 2017.
- 18) Natalia Ponomareva, Thomas Colthurst, Gilbert Hendry, Salem Haykal, and Soroush Radpour, Compact Multi-Class Boosted Trees, IEEE Big Data 2017.
- 19) . Arnab K. Paul, Arpit Goyal, Feiyi Wang, Sarp Oral, Ali R. Butt, Michael J. Brim, and Sangeetha B. Srinivasa, I/O Load Balancing for Big Data HPC Applications, IEEE Big Data 2017.
- 20) Shashank Gugnani, Xiaoyi Lu, Houliang Qi, Li Zha, and Dhabaleswar K. Panda, Characterizing and Accelerating Indexing Techniques on Distributed Ordered Tables, IEEE Big Data 2017.
- 21) Philipp Baumann, Dorit Hochbaum, and Quico Spaen, High-Performance Geometric Algorithms for Sparse Computation in Big Data Analytics, IEEE Big Data 2017.
- 22) . Yizhou Yan, Lei Cao, and Elke Rundensteiner, Distributed Top-N Local Outlier Detection in Big Data, IEEE Big Data 2017.
- 23) Chuxu Zhang, Lu Yu, Xiangliang Zhang, and Nitesh Chawla, ImWalkMF: Joint Matrix Factorization and Implicit Walk Integrative Learning for Recommendation, IEEE Big Data 2017.
- 24) Haekyu Park, Jinhong Jung, and U Kang, A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems, IEEE Big Data 2017.

- 25) Lorenzo De Stefani, Erisa Terolli, and Eli Upfal, Tiered Sampling: An Efficient Method for Approximate Counting Sparse Motifs in Massive Graph Streams, IEEE Big Data 2017.
- 26) Ankit Desai and Sanjay Chaudhary, Distributed Decision Tree v.2.0, IEEE Big Data 2017.
- 27) Byron Gao, Robert Tung, and Yong Yang, Iterative Matrix Correlation for Bisection Clustering, IEEE Big Data 2017.
- 28) Xinhui Tian, Yuanqing Guo, and Jianfeng Zhan, Towards Memory and Computation Efficient Graph Processing on Spark, IEEE Big Data 2017.
- 29) Celestine DÄnner, Thomas Parnell, Kubilay Atasu, Manolis Sifalakis, and Haralampos Pozidis, Understanding and Optimizing the Performance of Distributed Machine Learning: Applications on Apache Spark, IEEE Big Data 2017.
- 30) Lars Arge, Mathias Rav, Svend C. Svendsen, and Jakob Truelsen, External Memory Pipelining Made Easy With TPIE, IEEE Big Data 2017.
- 31) Hung Tran-The and Koji Zettsu, Discovering Co-occurrence Patterns of Heterogeneous Events from Unevenly-distributed Spatiotemporal Data, IEEE Big Data 2017.
- 32) Alexander Denzler and Michael Kaufmann, Toward Granular Knowledge Analytics for Data Intelligence, IEEE Big Data 2017.
- 33) Soukaina Filali Boubrahimi and Rafal Angryk, Time Series Classification Using a Neural Network Ensemble, IEEE Big Data 2018.
- 34) Rameshwar Pratap, Raghav Kulkarni, and Ishan Sohony, Efficient Dimensionality Reduction for Sparse Binary Data, IEEE Big Data 2018.
- 35) Yu-Min Chung, Chuan-Shen Hu, Austin Lawson, and Clifford Smyth, Topological approaches to skin disease image analysis, IEEE Big Data 2018.
- 36) Sebastian Werner, Jörn Kuhlenkamp, Markus Klems, Johannes Müller, and Stefan Tai, Serverless Big Data Processing using Matrix Multiplication as Example, IEEE Big Data 2018.
- 37) Vipul Gupta, Shusen Wang, Thomas Courtade, and Kannan Ramchandran, OverSketch: Approximate Matrix Multiplication for the Cloud, IEEE Big Data 2018.
- 38) Dianwei Han, Parallel DBSCAN Algorithm Using a Data Partitioning Strategy with Spark Implementation, IEEE Big Data 2018.
- 39) Daniel Rammer, Walid Budgaga, Thilina Buddhika, Shrideep Pallickara, and Sangmi Lee Pallickara, Alleviating I/O Inefficiencies to Enable Effective Model Training Over Voluminous, High-Dimensional Datasets, IEEE Big Data 2018.
- 40) Christian Beecks and Max Berrendorf, Optimal k-Nearest-Neighbor Query Processing via Multiple Lower Bound Approximations, IEEE Big Data 2018.
- 41) Atoshum Samuel Cahsai, Christos Anagnostopoulos, Nikos Ntarmos, and Peter Triantafillou, Revisiting Exact kNN Query Processing with Probabilistic Data Space Transformations, IEEE Big Data 2018.
- 42) Aparna Joshi, Yu Zhang, Petko Bogdanov, and Jeong-Hyon Hwang, An Efficient System for Subgraph Discovery, IEEE Big Data 2018.