

Algorithm Engineering for Big Data

Peter Sanders

sanders@kit.edu

Abstract: Perhaps the most fundamental challenge implied by advanced applications of big data sets is how to perform the vast amount of required computations sufficiently efficiently. Efficient algorithms are at the heart of this question. But how can we obtain innovative algorithmic solutions for demanding application problems with exploding input sizes using complex modern hardware and advanced algorithmic techniques? This tutorial gives examples how the methodology of algorithm engineering can be applied here. Examples include sorting, main memory based data bases, communication efficient algorithms, particle tracking at CERN LHC, 4D image processing, parallel graph algorithms, and full text indexing. Compared to a previous tutorial in Koblenz 2013 with the same title, this tutorial talks less about methodology and more about actual algorithms and applications. For further reading refer to [San13] and, for selected individual results to [DS03, KS07, SSP07, MS08, San09, RSS10, SS12, DS13].

Literatur

- [DS03] R. Dementiev und P. Sanders. Asynchronous Parallel Disk Sorting. In *15th ACM Symposium on Parallelism in Algorithms and Architectures*, Seiten 138–148, San Diego, 2003.
- [DS13] Jonathan Dees und Peter Sanders. Efficient Many-Core Query Execution in Main Memory Column-Stores. In *29th IEEE Conference on Data Engineering*, 2013.
- [KS07] F. Kulla und P. Sanders. Scalable Parallel Suffix Array Construction. *Parallel Computing*, 33:605–612, 2007. Special issue on Euro PVM/MPI 2006, distinguished paper.
- [MS08] K. Mehlhorn und P. Sanders. *Algorithms and Data Structures — The Basic Toolbox*. Springer, 2008.
- [RSS10] M. Rahn, P. Sanders und J. Singler. Scalable Distributed-Memory External Sorting. In *26th IEEE International Conference on Data Engineering*, Seiten 685–688, 2010.
- [San09] P. Sanders. Algorithm Engineering – An Attempt at a Definition. In *Efficient Algorithms*, Jgg. 5760 of *LNCS*, Seiten 321–340. Springer, 2009.
- [San13] Peter Sanders. Engineering Algorithms for Large Data Sets. In *39th Conf. on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, Jgg. 7741 of *LNCS*, Seiten 29–32. Springer, 2013. invited talk.
- [SS12] Peter Sanders und Christian Schulz. Distributed Evolutionary Graph Partitioning. In *ALENEX 2012*, Seiten 16–29. SIAM, 2012.
- [SSP07] J. Singler, P. Sanders und F. Putze. MCSTL: The Multi-core Standard Template Library. In *13th Euro-Par*, Jgg. 4641 of *LNCS*, Seiten 682–694. Springer, 2007.