ITMO University

Parallel algorithms for the analysis and synthesis of data

Course Workload		Assessment form (examination/ graded test/ ungraded test)
ECTS	Hours	
3	108	Ungraded test

This is an introductory course on parallel computing. During the course, students will learn the basics of parallel computing with the features of parallelizing various computer science tasks, as well as the advantages and disadvantages of technologies such as OpenMP, MPI and CUDA. This course describes various techniques for developing efficient parallel algorithms for a variety of tasks. We will also pay a lot of attention to the practical aspects of implementing parallel code, which really gives good performance in real parallel systems. Upon completion of the course, students will be able to apply their knowledge to solve complex computational problems. This course is designed for students who are interested in learning how to take advantage of parallel and distributed computing with an emphasis on writing parallel code.

Course structure:

1. Review of modern tools, technologies and tools for parallel programming: Multithreading, OpenMP, MPI, CUDA

1.1. Review of computing systems architectures. Classification of computing systems.
1.2. Modern technologies for the development of parallel programs: multithreading, OpenMP, MPI, CUDA.
1.3. Some ways to optimize parallel programs.

2. Data synthesis techniques: Pseudo-random number generation, parallel sorting algorithms

2.1. Overview of data synthesis methods.
2.2. Monte Carlo method and its significance in modeling.
2.3. Parallel algorithm for generating pseudo-random numbers.
2.4. Concepts for acceleration and efficiency of parallel algorithms.
2.5. Sorting algorithms: an overview of sequential sorting methods (bubble, insert, merge, hoare, swift, etc.) and their parallel counterparts.
2.6. Optimization methods for sorting algorithms.

3. DM technologies: classification and regression, clustering and association rules

3.1. Statement of the classification problem and presentation of the results. Classification rules: decision trees; mathematical functions. Methods for constructing classification rules.
3.2. Algorithm for constructing "1-rules"; method "Native Bayes". Algorithms for constructing decision trees. Linear methods. Least square method. Non-linear methods.
3.3. Statement of the problem of finding association rules. Varieties of tasks. Sequential analysis. Algorithm Apriori. Varieties of the Apriori algorithm.
3.4. Clustering problem statement. Distance types used in algorithms. Classification of clustering algorithms. Hierarchical algorithms. Non-hierarchical algorithms. Clustering quality and selection of the best solution. Adaptive clustering criteria. An example of adaptive clustering.

4. Analysis of text information: text mining

4.1. Tasks of text analysis. Stages of text analysis. Extraction of key concepts: the stage of local analysis. Extraction of key concepts: stage of integration and derivation of concepts.
4.2. Modern tools for text analysis. Information analysis algorithms

5. Distributed data analysis

5.1. Mobile agent systems. Using mobile agents for data analysis. Distributed data analysis system.