INSTITUTE OF COMPUTATIONAL MATHEMATICS
AND MATHEMATICAL GEOPHYSICS
SIBERIAN BRANCH OF RUSSIAN ACADEMY OF SCIENCES

# International conference
# ADVANCED MATHEMATICS, COMPUTATIONS AND APPLICATIONS – 2014

## ABSTRACTS

June 8-11, 2014
Akademgorodok, Novosibirsk, Russia

# Section 6. PROGRAMMING

## Large data clustering based on MapReduce Hadoop

*D. Akhmed-Zaki, M. Mansurova, A. Shomanov*

The goal of this study is the implementation of parallel algorithms for clustering of hyperspec-tral images. In this study parallel clustering algorithms are based on MapReduce programming mode and are implemented on the platform Hadoop. Research results are compared with the results with dif-ferent cluster settings and MPI model.

We have designed an experiment to gain data on working implementations of three Hadoop solutions for clustering problem: ISODATA without combiners, ISODATA with combiners and clus-tering using particle swarm optimization. The increasing of productivity is achieved by the process of data control. In this work the processed data are divided into two types: static, which are stored in the local file system, and states data that participate in the exchange between the Map and Reduce func-tions. Iterative algorithms use schemer that reduce the amount of data participating in the exchange of MapReduce processes. The results presented in this paper prove the effectiveness of MapReduce Hadoop technology for Remote Sensing data processing.

## Specialized multiagent platform for semantic text processing

*S.A. Anokhin, N.O. Garanina, E.A. Sidorova*

One of the fundamental problems being solved within the framework of intellectual systems is automation of knowledge acquisition from heterogeneous sources, in particular from natural language texts. Under the proposed approach information is retrieved from a textual documents in the form of facts expressed in terms of a given subject area model. Information represented in such a way then could be stored and interpreted easily by an intellectual system, in which all knowledge and data are formally described by a domain-specific ontology. One of the key features of the proposed conception is an application of a multiagent paradigm for describing a process of facts extraction from a textual data.

The proposed agent model implemented by specialized multiagent platform defines agents of two basic kinds: information agents corresponding to some entities of a subject domain recognized in a text, and rule agents corresponding to patterns of fact representation that perform computation of requested characteristics of information agents. The platform possess sufficient flexibility within the proposed framework and provides a way to effectively parallelize computation.

## Approaches addressing the problem of integration of ontology-based knowledge bases in the context of Linked Open Data

*Z.V. Apanovich, A.G. Marchuk*

Due to the fast progress of Semantic Web and its rapidly developing branch, Linked Open Data, large amounts of structured information from various fields are becoming available on the Web. Nowadays the Web of Data contains more than 28 billion RDF triples. New applications arise intended to use information from different data sources. This paper describes some problems of ontology alignment and identity resolution arising when employing the Linked Open Data datasets to enrich the content of a scientific knowledge base. The dataset of the Open Archive of the Russian Academy of Sciences, based on the BONE ontology, as well as various bibliographic datasets, structured by the AKT Reference ontology, are used as test examples. A template for SPARQL queries that establishes correspondence between groups of classes and relations of two ontologies is demonstrated. This template makes possible automatic generation of SPARQL queries based on the two ontologies visualization. The problem of identity resolution for bilingual identifiers of entities is discussed, and several approaches to its solution are proposed. All the experiments have been conducted with a toolkit developed at the IIS SB RAS. The experiments have shown that the datasets of the LOD cloud can be useful for the enrichment of the SB RAS Open Archive. In its turn, the content of the SB RAS Open Archive can be used for the refinement of information about researchers stored by many scholarly data repositories.