

SYSTEM OF HPC CONTENT ARCHIVING

*A. Bogdanov*¹, *A. Ivashchenko*²

St. Petersburg State University, St. Petersburg, Russia

This work is aimed to develop a system that will effectively solve the problem of storing and analyzing files containing text data, by using modern software development tools, techniques and approaches. The main challenge of storing a large number of text documents defined at the problem formulation stage has to be resolved with such functionality as full text search and document clustering depending on their contents. Main system features could be described with notions of distributed multilevel architecture, flexibility and interchangeability of components, achieved through the standard functionality incapsulation in independent executable modules.

Описана разработка системы, которая эффективно решает проблему хранения и анализа файлов, содержащих текстовые данные, с использованием современных программных инструментов, методов и подходов. Основная задача системы хранения большого количества текстовых документов состоит в предоставлении такой функциональности, как полнотекстовый поиск и кластеризация документов, в зависимости от их содержания. Основные функции системы описаны с помощью понятий распределенной многоуровневой архитектуры, гибкости и взаимозаменяемости компонентов, что достигается за счет инкапсуляции стандартной функциональности в независимых исполняемых модулях.

PACS: 07.05.Wr

Received on December 9, 2016.

¹E-mail: bogdanov@csa.ru

²aiivashchenko@cc.spbu.ru