

СИСТЕМА ВИРТУАЛЬНОГО ЛАБОРАТОРНОГО ПРАКТИКУМА ПО ПАРАЛЛЕЛЬНЫМ АЛГОРИТМАМ

И. В. Григорьева, Ю. В. Савицкий

THE SYSTEM OF VIRTUAL LABORATORY PRACTICE ON PARALLEL ALGORITHMS

I. V. Grigorieva, Yu. V. Savitskiy

Представленная работа описывает систему виртуального лабораторного практикума по параллельным вычислениям. Данная система обеспечивает доступ пользователей к теоретическому материалу, а также к серверам параллельных вычислений Центра коллективного пользования КемГУ.

The present work describes a system of virtual laboratory practice on parallel computing. This system allows the user access to the theoretical material, as well as parallel computing servers of Kemerovo State University Common Use Center.

Ключевые слова: виртуальная лаборатория, параллельные вычисления, параллельные алгоритмы.

Keywords: virtual laboratory, parallel computing, parallel algorithms.

Введение

В настоящий момент системы электронного образования (или E-learning) развиваются стремительными темпами. Большинство зарубежных высших образовательных учреждений уже имеют развитые информационные системы, такие например как www.coursera.org или www.udacity.com, и доступ к ним могут получить не только студенты, но просто люди со всего мира, желающие расширить свои познания.

Одним из видов электронных образовательных систем, который набирает популярность, является виртуальная учебная лаборатория. Определить термин «виртуальная учебная лаборатория» можно как программно-аппаратный комплекс, построенный с целью предоставления возможности дистанционного доступа к какому-либо редкому или ценному оборудованию или программному комплексу, либо к эмулятору данного комплекса, для проведения экспериментов [8].

В настоящий момент, комплексы подобного типа для различных областей знаний представлены ведущими российскими учебными заведениями, в качестве примеров можно привести: «ПараЛаб» – система Нижегородского государственного университета Н. И. Лобачевского [4], единый портал лабораторий удаленного доступа национальной нанотехнологической сети России [6], «Виртуальная образовательная лаборатория» [2] – виртуальная лаборатория, объединяющая в себе различные дисциплины: физику, химию, биологию, экологию.

Особое место системы «виртуальная лаборатория» занимают в области инженерного и информационного образования, поскольку они могут содержать в себе не только отдельные виртуальные опыты или макеты приборов, но и системы математического и имитационного моделирования, учебные и промышленные пакеты прикладных программ [4].

С другой стороны, параллельные вычисления являются одной из наиболее актуальных и перспективных отраслей информационных технологий. Параллельные технологии обработки данных повсеместно используются в самых передовых научных отраслях, таких как метеорология, геномная инженерия, фармацевтика, разработка наноматериалов и многих других,

которые требуют огромных вычислительных ресурсов [3; 7]. И, несмотря на то, что в настоящее время параллельные вычислительные технологии уже доступны широкому кругу пользователей, большинство из них не рассчитано на решение задач заявленной сложности. Вычислительные системы данного класса могут позволить себе владеть лишь крупные компании и вузы. И хотя нередко они предлагают предоставить свои ресурсы сторонним пользователям, но при этом стоимость и условия доступа часто являются неприемлемыми. Кроме того, системы параллельных вычислений имеют различный интерфейс, который также требует освоения.

Практика показала, что параллельное программирование существенно сложнее последовательного, поскольку в случае использования параллельных вычислительных устройств вычисления становятся недетерминированными, что существенно осложняет их тестирование и отладку. Построение и программная реализация алгоритмов для решения задач с использованием упомянутых систем требует приобретения компетенций в области теории параллельного программирования, освоения инструментов параллельного программирования и достаточной практики по реализации параллельных алгоритмов. Курсы по параллельному программированию входят в направления подготовки специальностей связанных с информатикой, компьютерными науками, математическим обеспечением, а также прикладной математикой. Система «Виртуальная лаборатория» была реализована для поддержки курсов «Параллельные алгоритмы», которые читаются по направлениям подготовки Фундаментальная информатика и информационные технологии, Прикладная математика и информатика и Математическое обеспечение и администрирование информационных систем математического факультета КемГУ.

Представленная в данной работе система «Виртуальная лаборатория» является новым вариантом системы удалённого доступа и управления распределёнными вычислительными ресурсами в рамках портала icr.kemsu.ru [5], основной целью которой было обеспечение и упорядочение доступа к параллельным вычислительным ресурсам Центра коллективного поль-

зования (ЦКП) высокопроизводительными вычислительными ресурсами КемГУ. Система «Виртуальная лаборатория» является, в первую очередь, образовательным ресурсом, предоставляющим теоретический материал и экспериментальную площадку для реализации параллельных алгоритмов в рамках одной системы. Система представляет собой web-приложение, которое реализует единый порядок удаленного доступа к различным вычислительным ресурсам. Другое назначение лаборатории – организация взаимодействия преподавателя и студентов, осваивающих курс «Параллельные алгоритмы».

Система используется для разработки параллельных программ на языке C/C++ для систем с распределенной памятью с использованием библиотеки MPI, систем с общей памятью с использованием директив OpenMP, а так же для гибридных систем с использованием MPI и OpenMP. Она предоставляет доступ к серверам Центра коллективного пользования по высокопроизводительным вычислениям КемГУ [1] master.kemsu.ru и учебный кластер «Полигон» (hpchead.kemsu.ru). Первый из перечисленных вычислительных ресурсов был получен по результатам конкурса, проведенного в 2009 году компанией Hewlett-Packard и национальным оператором связи «Синтерра» совместно с учреждениями Российской академии наук (Институт системного программирования РАН и Межведомственный суперкомпьютерный центр РАН) в рамках программы «Университетский кластер». На сегодняшний день вычислительный кластер состоит из 9 вычислительных узлов на основе блейд-шасси HP BL220c G5 (4 шт.) и HP BL280c G6 (5 шт.) и одного управляющего HP BL260c G5. Пиковая производительность составляет 1.5 Tflop/s. По протоколу iSCSI

к нему подключено сетевое хранилище NetApp FAS 3020 (размер раздела 1 Тб). В качестве операционной системы установлена Linux CentOS (5.5), а также следующие коммуникационные библиотеки: Intel MPI, MPICH 1.2.7, Open MPI 1.2.4, MPI LAM 7.1.4, UPC 2.10.0; компиляторы Intel C Compiler 11.0, Intel Fortran 11.0, gcc compiler 4.1.2; а также система пакетной обработки Open Torque 2.4.11.

Кластер «Полигон» является учебным кластером много меньшей мощности. В качестве операционной системы установлена та же Linux CentOS 5.5 и такой же набор компиляторов и библиотек. Немаловажным отличием от первого кластера является отсутствие системы пакетной обработки.

Приложение «Виртуальная лаборатория» позволяет использовать оборудование ЦКП в режиме обучения или выполнения научно-исследовательских работ широкому кругу обучающихся и сотрудников удаленно и используя при этом единый интерфейс доступа ко всем вычислительным ресурсам.

Другим существенным отличием представленного проекта является архитектурное решение, отличное от реализованного ранее в [5] (рис. 1). В системе [5] обмен данными между web-сервером и вычислительными ресурсами осуществлялся через базу данных. Web-сервер заполнял таблицы базы данных, хранящие файлы исходного кода программ и данные о том, на каком вычислительном ресурсе их необходимо разместить. Для распределения программ по вычислительным ресурсам был реализован дополнительный программный компонент – Менеджер вычислительных ресурсов, который также осуществлял сбор результатов выполнения программ и размещал их в базе данных, после чего web-сервер получал к ним доступ.

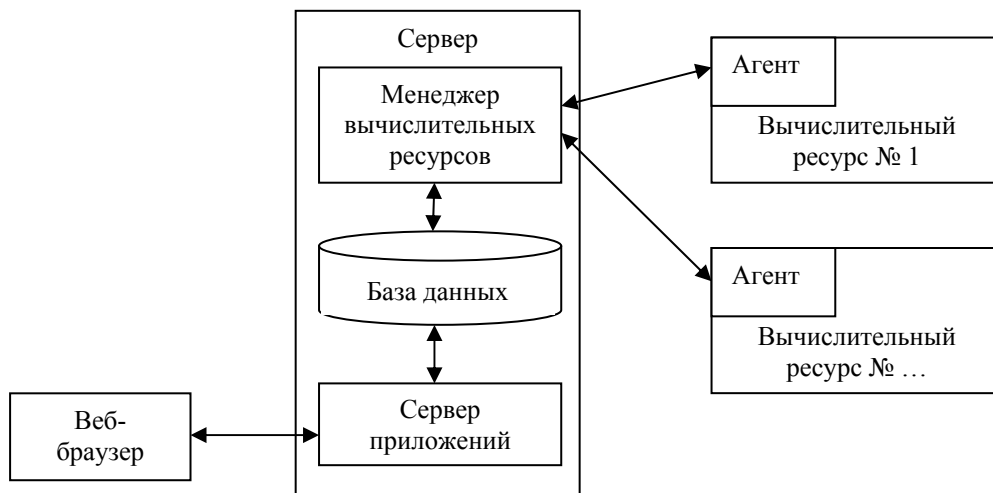


Рис. 1. Архитектура системы icp.kemsu.ru

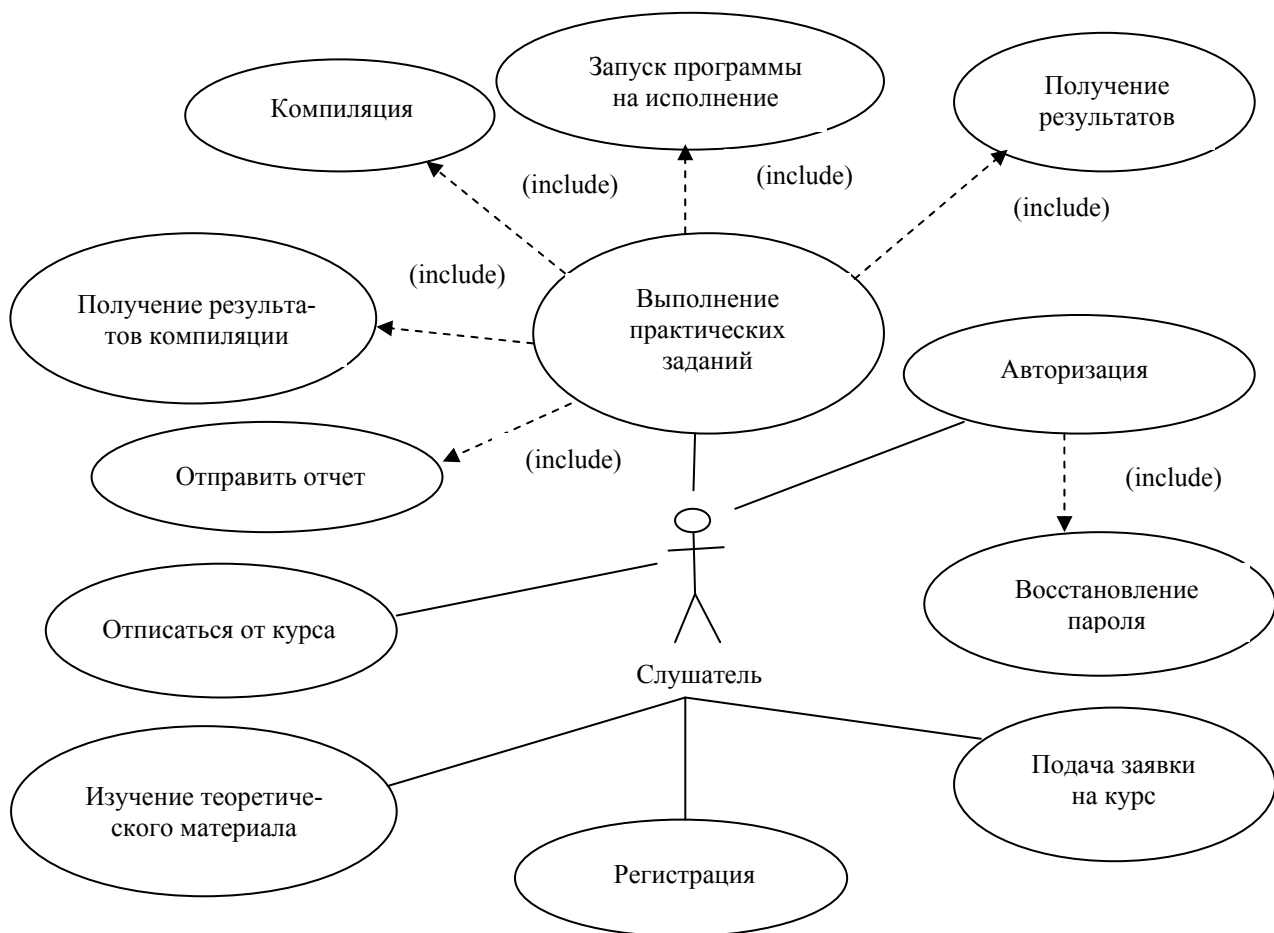


Рис. 2. Диаграмма вариантов использования роли Слушатель

В системе «Виртуальная лаборатория» web-приложение обращается к вычислительным ресурсам напрямую используя `remoteshell`, таким образом архитектура системы становится проще, кроме того снижается нагрузка на базу данных, так как для хранения файлов в нашем случае используется структура каталогов на диске.

Роли

В системе предполагается наличие пользователей трех типов: администратор, преподаватель и слушатель. Администратор осуществляет общее управление системой: управляет пользователями всех типов, редактирует (изменяет, актуализирует) материалы лабораторных работ и задания. Преподаватель выдает задания для выполнения лабораторных работ, просматривает, оценивает или отправляет на доработку отчеты по лабораторным работам, а также просматривает, компилирует и запускает программы слушателей для перепроверки результатов вычислений.

Большинство пользователей системы являются слушателями, поэтому эта роль рассматривается подробнее других. Диаграмма вариантов использования роли Слушатель представлена на рисунке 2. Слушатель имеет доступ к теоретическому материалу, и заданиям, назначенным преподавателем, реализует программы, осуществляет их компиляцию и генерирует выполняемый код программ на заданном вычисли-

тельном ресурсе, предварительно выбрав способ параллелизации (с использованием MPI, OpenMP или MPI и OpenMP одновременно) и компилятор. Если же программа содержит ошибки, то сообщения об ошибках возвращаются слушателю, после их исправления он может повторить попытку компиляции. Слушатель запускает выполняемый файл с заданными параметрами: количество процессоров и количество нитей, максимальное время выполнения, а также пользовательские параметры. Файл результатов, формирующийся в результате выполнения программы, также возвращается слушателю. После того, как все задания к данной лабораторной работе выполнены, слушатель отправляет отчет о выполнении лабораторной работы преподавателю.

Архитектура системы

Архитектура системы представлена на рисунке 3. Система включает в себя следующие компоненты: web-агент – клиентское приложение, в качестве которого выступает web-браузер, web-сервер формирует страницы, выполняя запросы к базе данных. База данных обрабатывает информацию о страницах, пользователях, лабораторных работах и заданиях, вычислительных ресурсах, доступных действиях и командах, осуществляет учет файлов исходного кода, результатов компиляции, выполняемых файлов, файлов результатов. Все перечисленные файлы, за исключени-

ем исполняемых файлов, хранятся в структуре каталогов на web-сервере, обозначенном на рисунке 3 как Хранилище файлов. Web-сервер пересылает файлы исходного кода на вычислительные ресурсы и сохраняет их в структуре каталогов. Файлы исходного кода отправляются на компиляцию, результатом которой является выполняемый файл или файл с сообщениями

об ошибках, который возвращается web-серверу. Исполняемые файлы сохраняются в каталогах вычислительного агента. Web-сервер осуществляет запуск программ на выполнение с помощью remoteshell согласно регламента вычислительного ресурса, на котором осуществляется запуск, и с заданным набором параметров.

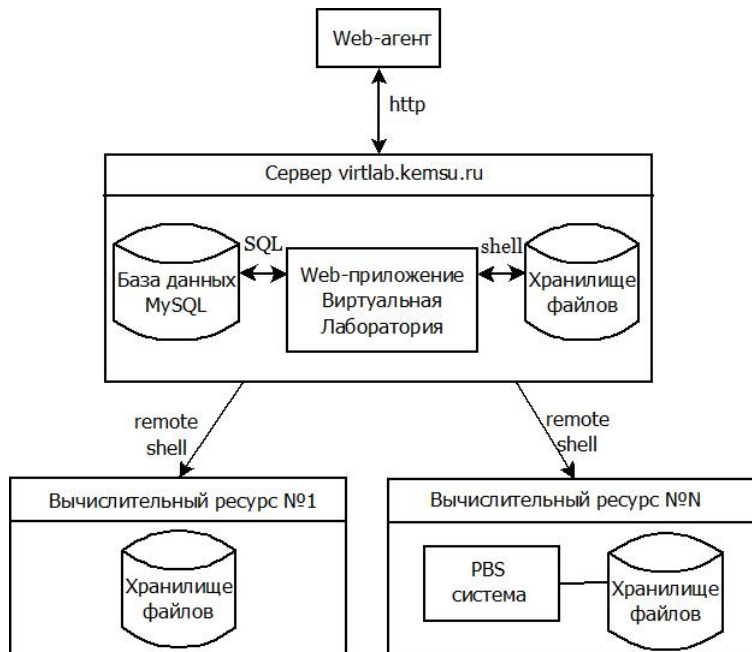


Рис. 3. Архитектура системы

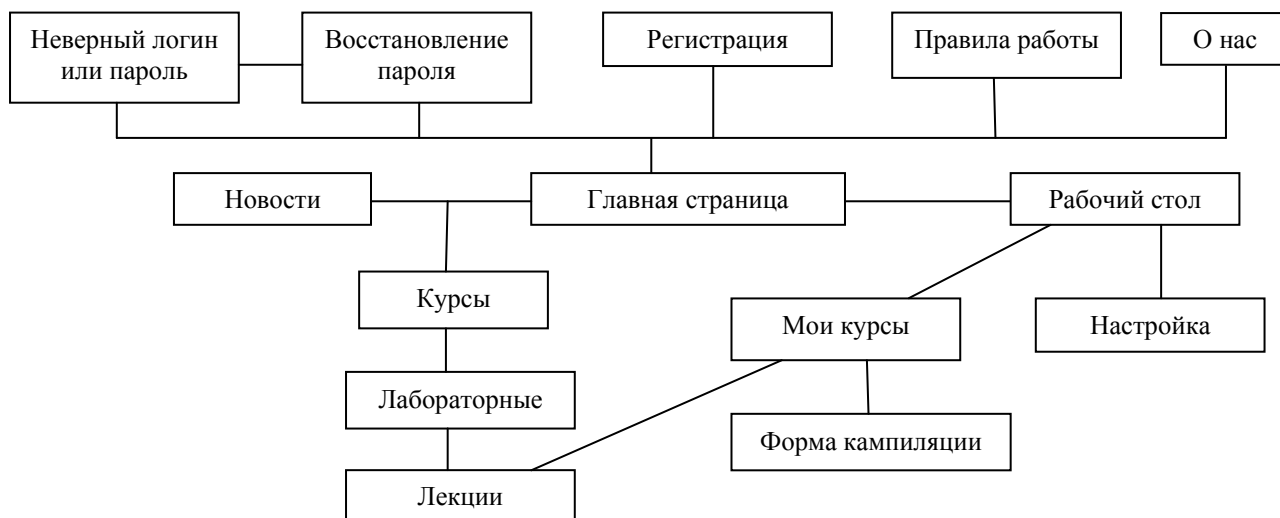


Рис. 4. Карта сайта

Структура страниц системы представлена на рисунке 4. Все выполняемые слушателем в настоящее время задания доступны ему из рабочего стола. Важным преимуществом представленной системы по сравнению с системой [5] является наличие страницы онлайн-компиляции, в которой слушатель может работать с кодом программы. Код подсвечивается, так же, как и в стандартных средах разработчика. Код программ с этой страницы отправляется web-сервером на компиляцию, ошибки компиляции, если такие

были найдены, отображаются на странице. После исправления ошибок слушатель осуществляет компиляцию повторно.

В настоящее время система находится в тестовой эксплуатации, используется при чтении дисциплины «Параллельные алгоритмы» на математическом факультете КемГУ. Интерфейсы ролей Администратор и Преподаватель находятся на стадии тестирования. Интерфейс роли Администратор совпадает с интерфейсом администратора CMS/CMF-системы ModX, с

использованием которой реализована данная система, которая дополнена набором вкладок для управления студентами, курсами, заданиями и вычислительными ресурсами.

Средства реализации

В качестве средств реализации были выбраны:

1. CMS/CMF MODx – система управления контентом, призвана облегчить разработку web-приложения. Данная система использует язык PHP, исполняющий на стороне сервера Apache. В данном случае CMS система используется с тем расчётом, что она возьмёт на себя часть функций ролей администратора и преподавателя. Основными функциональными элементами системы являются: ресурсы – являющиеся страницами сайта, шаблоны – задают стиль страниц, чанки – элементы, содержащиеся в себе html код, а также сниппеты – элементы, содержащие PHP код, данные элементы могут быть вызваны из чанков или из ресурсов MODx.

2. Apache был выбран в качестве web-сервера по той причине, что он является свободно распространяемым продуктом и, вместе с тем, является достаточно надёжным и гибким программным средством.

3. В качестве базы данных была выбрана СУБД MySQL, которая также является свободно распространяемой.

4. Torque – менеджер распределенных вычислительных ресурсов для кластеров под управлением Linux и других Unix-подобных операционных систем, называемый также PBS-система (Portable Batch System). При одновременном выполнении заданий несколькими пользователями на одном вычислительном комплексе Torque позволяет оптимально использовать вычислительные ресурсы: свести к минимуму как перегрузку какого-либо одного узла, так и его простой.

5. Maui – планировщик заданий в параллельных и распределённых вычислительных системах, используется взамен штатных планировщиков.

Выборанный подход к реализации системы и набор средств реализации позволяет утверждать, что систе-

ма «Виртуальная лаборатория» удовлетворяет стандарту открытых информационных систем [9].

Заключение

В результате проведенной работы была реализована система «Виртуальная лаборатория», апробирован новый подход к построению архитектуры системы. Система находится в тестовой эксплуатации и используется при чтении спецкурса «Параллельные алгоритмы» на математическом факультете КемГУ. В результате апробации выбранный подход к построению архитектуры доказал свою жизнеспособность. Участвовавшие в апробации студенческие группы отметили положительные стороны использования «Виртуальной лаборатории»: повышение доступности вычислительных ресурсов, которые ранее были доступны только из локальной сети университета или с использованием VPN-соединения, наличие полного теоретического материала, удобство работы со страницей компиляции, а также возможность взаимодействия с преподавателем.

Одной из особенностей системы «Виртуальная лаборатория» является возможность добавления и поддержки новых курсов с доступом к другим программно-аппаратным комплексам без переработки кода приложения с использованием страниц добавления и редактирования курсов в интерфейсе администратора. В результате дополнения системы новыми курсами будет получена система виртуальных лабораторных практикумов по ряду дисциплин.

Другой особенностью является ее соответствие стандарту открытых информационных систем [9]. Применение подхода открытых систем в настоящее время является основной тенденцией в области информационных технологий. Результативность этого подхода определяется реализацией стандартов открытых систем [9], которые специфицируют интерфейсы, службы и поддерживаемые форматы для обеспечения интероперабельности, масштабируемости и мобильности приложений.

Литература

1. Афанасьев, К. Е. Предпосылки создания центра распределенных высокопроизводительных вычислений в КемГУ / К. Е. Афанасьев, А. М. Гудов, А. В. Демидов; под ред. К. Е. Афанасьева // Информационные и вычислительные технологии в численных расчетах и управлении вузом: сб. ст. – Кемерово, 2010. – С. 471 – 477.
2. Виртуальная образовательная лаборатория. – Режим доступа: // (<http://www.virtulab.net/>)
3. Воеводин, В. В. Параллельные вычисления / В. В. Воеводин, Вл. В. Воеводин. – БХВ-Петербург, 2004. – 608 с.
4. Гергель, В. П. Паралаб. Программная система для изучения и исследования методов параллельных вычислений: учеб. пособие / В. П. Гергель, А. А. Лабутина. – Нижний Новгород: ННГУ им. Н. И. Лобачевского, 2003. – 125 с. – Режим доступа: // (<http://www.software.unn.ru/ccam/files/ParaLab/ParaLabTutorial.pdf>)
5. Григорьева, И. В. Система удаленного доступа и управления распределенными вычислительными ресурсами / И. В. Григорьева, А. В. Демидов // Вычислительные технологии. – 2008. – Т. 13. – С. 28 – 32.
6. Единый портал лабораторий удаленного доступа национальной нанотехнологической сети России. – Режим доступа: <http://www.nano-network.ru>
7. Суперкомпьютерное моделирование в физике климатической системы: учебное пособие / В. Н. Лыков [и др.]. – М.: Издательство Московского гос. университета, 2012. – 408 с.
8. Соловов, А. В. Электронное обучение: проблематика, дидактика, технология / А. В. Соловов. – Самара: Новая техника, 2006. – 464 с.
9. W3C Web Svices Architecture / W3C. – W3C Working Group Note 11.02.04. - Электрон. текстовые дан. – [Б. м]: W 3 C, 2004. – Режим доступа: <http://www.w3.org/TR/ws-gloss/ws-arch.html>

Информация об авторах:

Григорьева Ирина Владимировна – научный руководитель, кандидат физико-математических наук, доцент математического факультета кафедры ЮНЕСКО по НИТ КемГУ, 546469, 8-950-580-76-93, igriva@list.ru.

Irina V. Grigorieva – research advisor, Candidate of Physics and Mathematics, Associate Professor, Assistant Professor at the UNESCO Chair for New Information Technologies, Kemerovo State University.

Савицкий Юрий Васильевич – бакалавр информационных технологий КемГУ, 8-923-460-69-806, mr-l-ik@yandex.ru.

Yuriy V. Savitskiy – Bachelor in Information Technologies, Kemerovo State University.