



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛМ ЖӘНЕ ФЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҮЛТТЫҚ УНИВЕРСИТЕТІ



СОВЕТ МОЛОДЫХ УЧЕНЫХ
Еуразийский национальный университет им.Л.Н.Гумилева

**Студенттер мен жас ғалымдардың
«ФЫЛЫМ ЖӘНЕ БІЛМ - 2014» атты
IX халықаралық ғылыми конференциясы**

**IX Международная научная конференция
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ - 2014»**

**The IX International Scientific Conference for
students and young scholars
«SCIENCE AND EDUCATION-2014»**

2014 жыл 11 сәуір
11 апреля 2014 года
April 11, 2014



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҮЛТТЫҚ УНИВЕРСИТЕТІ

**Студенттер мен жас ғалымдардың
«Ғылым және білім - 2014»
атты IX Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
IX Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2014»**

**PROCEEDINGS
of the IX International Scientific Conference
for students and young scholars
«Science and education - 2014»**

2014 жыл 11 сәуір

Астана

УДК 001(063)

ББК 72

F 96

F 96

«Ғылым және білім – 2014» атты студенттер мен жас ғалымдардың IX Халықаралық ғылыми конференциясы = IX Международная научная конференция студентов и молодых ученых «Наука и образование - 2014» = The IX International Scientific Conference for students and young scholars «Science and education - 2014». – Астана: <http://www.enu.kz/ru/nauka/nauka-i-obrazovanie/>, 2014. – 5831 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-610-4

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 001(063)

ББК 72

ISBN 978-9965-31-610-4

©Л.Н. Гумилев атындағы Еуразия ұлттық университеті, 2014

2. Балапанов Е, Бөрібаев Б, Бекбаев А, Дәuletқұлов А, Спанқұлова Л, Информатика терминдерінің қазақша-ағылшынша-орысша орысша-қазақша- ағылшынша-орысша-қазақша сөздігі.
3. Крейгlin X. «Архитектура компьютеров и ее реализация». «Мир», 2004
4. Михаил Буза Архитектура компьютеров. Учебник. «Новое знание», 2006
5. Мураховский В.И. Сборка, настройка, апгрейд современного компьютера. «Десском», 2000.
6. Невзоров В. Как самому собрать компьютер, устройство компьютера, оптимальная конфигурация, руководство по сборке. «Питер», 2000

УДК 004.7

ПЕРСПЕКТИВА ИСПОЛЬЗОВАНИЯ СИСТЕМЫ HADOOP ДЛЯ АНАЛИЗА ДАННЫХ В WEB-СТРАНИЦАХ

Утепова Шолпан Бахитовна

shopik_7@mail.ru

Магистрант факультета информационных технологий

ЕНУ им. Л.Н.Гумилева, Астана, Казахстан

Научный руководитель – М. Серік

Современный период характеризуется как период бурного развития информатизации. В информационном обществе образование является основным приоритетным направлением и основным индикатором развития. Таким образом, повышение качества подготовки специалистов в ВУЗах является приоритетным.

Hadoop - это каркас с открытым исходным кодом, предназначенный для создания и запуска распределенных приложений, обрабатывающих большие объемы данных и является актуальной разработкой определенных методологических подходов к использованию информационно-поисковых технологий в учебном процессе.

Цель исследования заключается в повышении эффективности поиска, извлечения и анализа проблемно-ориентированных естественно-языковых текстов, размещенных в Web-страницах, за счет использования передовых технологий распределенных вычислений на базе Hadoop.

Методология масштабирования простой программы для обработки большого набора данных.

Рассмотрим количество одинаковых слов в наборе документов, если оно содержит в себе единственное предложение «Делай, как я говорю, а не как я делаю». Мы получим счётчик слов, с помощью простой программы define wordCount as Multiset;

```
for each document in documentSet {  
    T = tokenize(document);  
    for each token in T {  
        wordCount[token]++; } }  
    display(wordCount);
```

Каждый документ разбивается на слова, и для каждого слова на единицу увеличивается соответствующий ему счетчик в мульти множестве wordCount. В конце функция display() распечатывает все элементы wordCount [1].

Для простой программы количество набора документов ограничено. Перебор всех этих документов на одном компьютере занимает чрезвычайно много времени. Распределение работы между несколькими машинами может сократить время. В таком случае псевдокод данной программы будет содержать мульти множество totalWordCount хранящийся в памяти. Тут возникает другая проблема, посягание на доступную оперативную память. Так как структура wordCount может не поместиться в память, надо переписать программу так, чтобы

данная хеш-таблица хранилась на диске. Реализация дисковой хеш-таблицы требует достаточно сложный и объемный код. Использование каркаса Hadoop решает данную проблему на основе модели MapReduce. С помощью MapReduce можно осуществить эффективную фильтрацию, а затем и преобразовать входные данные. Именно поэтому MapReduce является основой каркаса Hadoop, что упрощает возможность масштабирования программ высоконагруженной обработки больших объемов данных.

Существуют крупные системы, построенные на базе Hadoop, – это системы, которые организуют кластер Hadoop на собственной площадке, для решения реальных практических задач. Система Hadoop позволяет использовать технологию обхода и индексирования Web-страниц. Что в перспективе дает возможность интеллектуального анализа социологических данных, которые используются в борьбе с экстремизмом, терроризмом и облегчает работу службы государственной безопасности, что является на сегодняшний день актуальным.

В образовательной среде система Hadoop обеспечивает возможность внедрения в учебный процесс информационно-поисковой технологии обучения, повышения эффективности поиска, извлечения и анализа проблемно-ориентированных естественноязыковых текстов, размещенных в Web-страницах.

Вот как используют Hadoop известные и не очень известные компании: Adobe применяет Hadoop для социальных сервисов, структурированного хранения данных и внутренней обработки информации. Сейчас работает 30-узловой кластер и планируется развертывание 80-узлового; Brockmann Consult занимается вопросами обработки данных о состоянии окружающей среды и геоинформационными сервисами. Компания использует параллельную обработку больших объемов информации, получаемых со спутников, с помощью своей системы Calvalus, базирующейся на Hadoop; в eBay в кластере из 532 узлов Hadoop служит для оптимизации поиска и исследования данных; Eyealike применяет Hadoop в своей технологии поиска изображений и видео для таких задач, как поиск похожих лиц или автоматическое проставление тегов в социальных медиа; Facebook располагает двумя большими кластерами Hadoop для хранения внутренних журналов и источников данных для создания отчетов, аналитики и машинного обучения; Hulu использует Hadoop для хранения и анализа журналов на кластере из 13 компьютеров; Kalooga, сервис поиска в галереях изображений, применяет 20-узловой кластер Hadoop для веб-сканирования, анализа, и обработки событий; Twitter использует Hadoop для хранения и обработки всех твитов и других типов данных, сгенерированных в социальной сети; Yahoo располагает кластером Hadoop из 4500 узлов для исследования данных своих веб-серверов, и в рекламной системе. Он также служит для тестов масштабирования при разработке Hadoop для больших кластеров [2].

Система Hadoop имеет следующие возможности: использование для интеллектуального анализа социологических данных; выявление теоретических основ технологии разработки эффективных и распределенных методов обхода и индексирования Web-страниц; реализация методики на примере технологии автоматического извлечения знаний из проблемно-ориентированных естественноязыковых текстов; разработаны модель и содержание информационно-поисковой технологии обучения, формирующая фундаментальные и универсальные знания, умения и навыки, которые позволяют будущим специалистам адаптироваться в их профессиональной деятельности;

С каждым днём в мире происходят открытия в той или иной области. Наша профессиональная деятельности включает в себя не только знания информатики, а обучения ему учащихся средних и высших учебных заведений. Обучая информатике нужно быть не только образованным человеком, но и хорошим психологом. При этом знания, которые ты даешь, нужно всегда улучшать и самому заниматься повышением квалификации, следя за последними новшествами происходящие в информационной среде.

Учитывая на сегодня, что каждый может в этой области науки заниматься самостоятельно. Например, через всемирную сеть, что может резко воздействовать на авторитет специалиста, нужно быть всегда готовым к беседам о последних новостях в

области информатики. Это еще раз говорит о том, что нельзя останавливаться на достигнутом, подавно информатика как наука является молодой. И думаю бесконечной, как впрочем, и другие науки. Система, которую мы исследуем, возможно, будет применена как инструмент в решении задач в области информатики и программирования. [3,4]

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hadoop в действии. - М.: ДМК Пресс, 2012
2. <http://www.pcweek.ru/idea/article/detail.php?ID=136413>
3. Труды XIX Всероссийской научно-методической конференции «Телематика-2012». – Санкт-Петербург: Санкт-Петербург.нац. исслед.. унив.информ.техн., механики и оптики, 2012. –С.17-18.
4. Международный журнал экспериментального образования. Секция педагогических наук. - М.: Российская Академия Естествознания, 2013. -№1. –С. 98-100. РИНЦ 0,548.

УДК (004.42 + 004.65)

СОЗДАНИЕ АВТОМАТИЗИРОВАННОЙ КЛИЕНТ-СЕРВЕРНОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ

Утюганов Филипп Сергеевич, Тукубаев Эльдар Эдикович, Рыбальченко Андрей Сергеевич

philipp.u92@gmail.com, eldar_kokshe2009@mail.ru, smallwhitetiger@mail.ru

Студенты 4 курса кафедры информатики ЕНУ им. Л.Н. Гумилева, Астана, Казахстан
Научный руководитель – Е.Е. Омарбеков

С развитием информационных технологий широкое распространение получили различные системы тестирования. В настоящее время наблюдается интенсивный рост заинтересованности в таких системах, как со стороны учебных заведений, так и со стороны учащихся. Причиной такого интереса является потребность автоматизации большей части проверки знаний учащихся и возможность обучаться на расстоянии, что снижает нагрузку на преподавателей и повышает доступность обучения [1, 2].

В настоящей статье рассматривается автоматизированная клиент-серверная система тестирования. Основной проблемой при тестировании учащихся является необходимость автоматизации. Тесты должны содержать большое количество динамических иллюстраций «визуализаторов».

В последнее десятилетие в мире стране был разработан ряд дистанционных систем тестирования, обеспечивающих информационную, методическую и учебно-тренировочную поддержку проверки знаний учащихся. Основными их составляющими являются базы тестовых заданий и респондентов с результатами тестирования и статистикой, студия разработки тестов, интерфейс общения с тестовой системой, правила допуска и процедура прохождения тестирования, способы оценивания и обработки статистики. С точки зрения технологий программирования система тестирования знаний должна обладать платформенной независимостью, работать в сетевом варианте и использовать свободно распространяемые программные продукты. Рассматриваемая нами система тестирования построена на основе Web-технологий и может работать как в корпоративной сети (INTRANET), так и в глобальной сети INTERNET.

Описание концепции

В основе автоматизированной системы тестирования учащихся лежит разделение процесса тестирования на некоторое количество этапов.

Условно весь процесс тестирования можно разделить на несколько этапов:

- Авторизация учащегося
- Выбор предмета