

Используя интегрированную среду разработки программного обеспечения Delphi, авторы разработали программный продукт, позволяющий автоматизировать перерасчет учебной нагрузки преподавателей кафедры. Входными данными служат: дата изменения учебной нагрузки, «старый» и «новый» объемы учебной нагрузки (в соответствии с первоначальным и актуализированным учебными поручениями преподавателя).

Перераспределение учебной нагрузки производится равномерно на оставшийся период учебного года, начиная с указанной даты. При этом учитывается объем отработанных часов и вычисляется объем часов, подлежащих проведению. Программа в автоматическом режиме формирует следующие документы: заявление преподавателя на отмену/изменение учебной нагрузки, заявление преподавателя на совмещение научно-педагогической работы. Также предусмотрена возможность вывода различных вариантов отчетов.

Программа имеет интуитивно понятный интерфейс и позволяет уменьшить рутинную работу в результате автоматизации оформления документов и отчетов.

УДК 502.7 + 577.4

**Разработка компьютерной системы
обработки больших данных
для персонализированного анализа особенностей генома**

А.А. Цхай^{1,2}, С.В. Мурзинцев²

¹ АлтГТУ им. И.И.Ползунова, ² АлтГУ, г. Барнаул

Изучение структурно-функциональной организации биоразнообразия продолжает оставаться магистральным направлением, развивающимся на стыке биологии и информационных технологий [1].

В последнее время подход, основанный на сравнении геномов, стал весьма широко и эффективно использоваться в генетике и молекулярной биологии. Благодаря развитию математических алгоритмов и информационных технологий данный подход выделился в новое самостоятельное направление под названием «сравнительная геномика». Это направление объединяет исследования, посвященные сопоставлению свойств совокупностей генов различных живых организмов.

На данный момент времени исследователи, помимо широко известным результатов по геному человека, обладают расшифрованными геномами как простейших организмов (дрожжи, мухи дрозофилы, кишечная палочка), так и сложных живых существ, таких как обезьяны,

крысы, собаки и т.д. Параллельно ведется расшифровка ещё множества других геномов, как организмов, так и растений. Текст ДНК во всех геномах написан четырьмя буквами, у птиц число этих букв составляет сотни миллионов, у человека и млекопитающих «текст» ДНК доходит до миллиарда символов. Поскольку запись генома любого существа осуществляется четырьмя буквами, то возникает возможность сравнения геномов живых существ между собой для выявления закономерностей и поиска соответствия, например, с отклонениями или нарушениями в жизнедеятельности изучаемых организмов. В свою очередь это позволяет составлять БД организмов и БД отклонений, нарушений в ДНК. Это в перспективе позволит находить и анализировать отклонения в геномах живых существ, в том числе человека, на ранних стадиях развития и заблаговременно заниматься профилактикой и лечением.

В статье получены результаты, позволяющие найти подтверждение гипотезе о применимости распределенных нереляционных БД для сравнения и поиска различных поведенческих отклонений на основе персонифицированного анализа геномов живых существ.

Для поиска отклонений в геномах живых существ использованы алгоритм Wnnowing и распределенная система поиска и хранения данных MongoDB представленная на рисунке 1 [2–4].

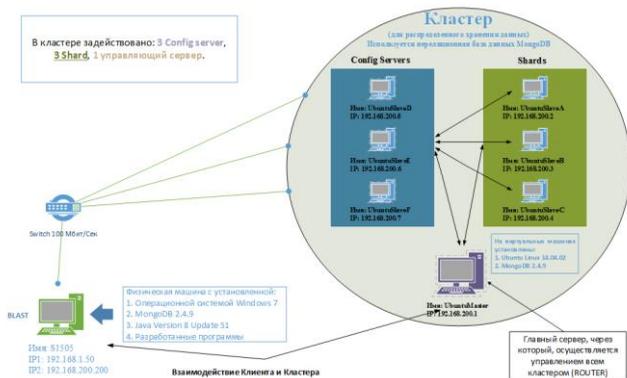


Рисунок 1 – Схема распределенной информационной системы для поиска отклонений у организмов

В данном исследовании в качестве источника элементов БД геномов организмов была выбрана БД Киотской энциклопедии генов и геномов KEGG GENOME (Секвенированные геномы различных живых

мами живых существ и времени загрузки отклонений и геномов организмов в БД.

Оценка масштабирования инфраструктуры, осуществлялась благодаря построению прототипа ИС с использованием различных конфигураций. Были протестированы конфигурации без разделения записей по индексу, а так же с разделением записи по индексу между одной шардой и тремя.

Продолжение исследований видится в нескольких направлениях. Первое направление – это решение задачи об определении момента, при котором, необходимо добавлять узел к кластеру при возрастании рассматриваемого количества отклонений и увеличении числа геномов в БД организмов.

Второе – это практическое наполнение и дальнейшее формирование БД с как можно большим количеством реальных геномов существ. Использование полученных результатов в междисциплинарных геномных системных исследованиях позволило бы говорить о детализации и развитии модели в перспективном плане.

Библиографический список

1. Биоразнообразие и динамика экосистем: информационные технологии и моделирование / отв. ред. В.К. Шумный, Ю.И. Шокин, Н.А. Колчанов, А.М. Федотов. - Новосибирск: Изд-во СО РАН, 2006. – 648 с.

2. Цхай А.А., Бутаков С.В., Мурзинцев С.В., Ким Л.С. Обнаружение плагиата с использованием нереляционных баз данных // Вестник алтайской науки.– 2015. – №1. – С. 280–285.

3. Бутаков С.В., Мурзинцев С.В., Цхай А.А. Использование горизонтально-масштабируемой инфраструктуры при поиске заимствований в тексте // Информационные технологии. – 2016. – Т. 22, №8. - С. 576–581.

4. Butakov S., Murzintsev S., Tskhai A. (2016). Detecting Text Similarity on a Scalable No-SQL Database Platform. Platform Technology and Service (PlatCon). Proc. of Int. Conf. 2016; doi: 10.1109/PlatCon.2016.7456789.