

«семантическое пространство», в котором и производятся все дальнейшие сравнения. В контексте поставленной задачи документом будет являться ответ, а слова – извлекаться из запроса пользователя. На первом шаге необходимо составить частотную матрицу индексируемых слов. В этой матрице строки соответствуют индексируемым словам, а столбцы – ответам. В каждой ячейке указано, какое количество раз слово встречается в соответствующем ответе. На данном шаге необходимо учитывать при подсчете слова-синонимы. Это позволит значительно улучшить исходный алгоритм, однако для этого требуется иметь в наличии словарь синонимов. На следующем этапе мы проводим сингулярное разложение полученной матрицы. Это математическая операция разложения матрицы на три составляющие:

$$M = U * W * V^T,$$

Где  $U$  и  $V^T$  – ортогональные матрицы, а  $W$  – диагональная. Диагональные элементы матрицы  $W$  отсортированы в порядке убывания. Диагональные элементы матрицы  $W$  называются сингулярными числами. Основное достоинство сингулярного разложения заключается в том, что оно позволяет выделить ключевые элементы матрицы и проигнорировать шумы. Поскольку меньшие сингулярные числа вносят незначительный вклад в итоговое произведение, размерность матрицы можно уменьшить. Единогласного решения в выборе результирующей размерности нет, т.к. при маленькой размерности снижается возможность обнаруживать семантические группы, а при большой размерности на результат начинают оказывать влияние шумы. После того как матрица была сформирована, необходимо найти наименьшее расстояние от набора слов запроса до ответов. Значениями слов для представления их в многомерном пространстве будут являться строки матрицы  $U$ , а ответов – столбцы матрицы  $V^T$ . Поскольку пространство отображения зачастую получается многомерным, выбор способа вычисления расстояния становится очень важным для производительности алгоритма. Среди всех вариантов было решено использовать манхэттенское расстояние, т.к. оно обладает высокой точностью и низкой сложностью алгоритма.

Применение данного алгоритма позволяет реализовать автоматического агента, который будет способен решить большое количество проблем, ответить на многие вопросы пользователей, уменьшить нагрузку на службу поддержки, в которой задействованы люди.

Список использованных источников:

1. Landauer T. K., Foltz P. W., Laham, D. Introduction to Latent Semantic Analysis // Discourse Processes. 1998. No 25. P. 259–284
2. Golub G. H., Luk F. T., Overton M. L. A block Lanczos method for computing the singular values and corresponding singular vectors of a matrix // ACM Transactions on Mathematical Software 1981. No 7. P. 149–169.

## ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ MAPREDUCE ДЛЯ ОРГАНИЗАЦИИ РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ СИГНАЛЬНЫХ ДАННЫХ

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Казимирчик Д.В.*

*Бранцевич П.Ю. – к. т. н., доцент*

Обработка больших массивов экспериментальных данных с использованием стандартных алгоритмов и методов может занимать достаточно продолжительное время. В таких случаях представляется целесообразным использование методов, которые позволяют ускорить обработку данных распределяя задачи обработки между множеством машин.

Сегодня стоимость хранения информации настолько низка, что зачастую представляется целесообразным постоянное накопление “сырых” экспериментальных данных получаемых от датчиков, систем мониторинга событий и т.д. в реальном времени. В будущем над собранными в процессе наблюдения данными можно проводить различные виды анализа для получения наиболее полной картины происходящего, выявления закономерностей и аномалий. В результате этого появляется необходимость в организации эффективной и быстрой обработки таких данных. Так как ресурс повышения производительности отдельно взятых процессоров давно исчерпан, постоянно ведётся поиск путей ускорения вычислительных процессов с использованием методов распараллеливания и распределения обработки данных между множеством машин.

MapReduce – это модель организации вычислений предназначенная для использования при обработке и генерации больших объёмов данных. При использовании MapReduce выбирается мар-функция, которая обрабатывает исходную пару параметров ключ/значение и в результате генерирует промежуточную пару ключ/значение, и reduce-функция, которая объединяет все промежуточные значения ассоциированные с одним промежуточным ключом [1].

Модель MapReduce может быть применена при обработке большого количества сигнальных данных для вычисления таких параметров сигнала, как среднее квадратическое значение (СКЗ), пик-фактор, размах колебаний, а также для проведения более сложных видов анализа и обработки. Базовым подходом при реализации MapReduce в этом случае является разбиение массива сигнальных значений на части, которые в дальнейшем будут отданы на обработку множеству независимо исполняемых мар-функций, которые

подсчитывают промежуточный результат для обрабатываемых ими групп данных. Затем, после завершения работы *map*-функций, *reduce*-функция объединяет результаты их работы вычисляя итоговое значение необходимого параметра.

Так для вычисления размаха колебаний, среднего квадратического значения и пик-фактора сигнала, *map*-функция может быть определена таким образом, чтобы вычислять минимальное и максимальное значение и сумму квадратов части массива дискретных значений сигнала. *Reduce*-функция затем получает эти промежуточные данные и заканчивает вычисления, находя максимальное и минимальное значения среди промежуточных, которые как очевидно и будут являться максимумом и минимумом среди всех значений массива данных, а также суммирует промежуточные суммы квадратов, из которых вычисляется среднеквадратичное значение, а затем и пик-фактор для сигнала.

Также данный подход может использоваться при выполнении цифровой фильтрации сигнала. В этом случае массив значений исходного сигнала делится на части, которые затем независимо обрабатываются *map*-функциями. *Map*-функция в данном случае выполняет прямое быстрое преобразование Фурье, удаляет ненужные частотные составляющие и производит обратное преобразование Фурье, полученный результат, являющийся отфильтрованной частью сигнала сохраняется как промежуточный результат для *reduce*-функции. *Reduce*-функция собирает части сигнала обработанные *map*-функциями в один цельный результирующий сигнал.

На основании вышеизложенного было создано программное средство, с использованием *MapReduce* функциональности, предоставляемой фреймворком *ApacheHadoop* [2], и распределённой файловой системы *HDFS* [3], которое при использовании нескольких машин позволило ускорить обработку большого объёма сигнальных данных по сравнению с традиционной обработкой на одной машине.

Таким образом, технология *MapReduce* при её применении с подходящими алгоритмами обработки данных может дать значительный прирост в скорости обработки данных путём распределения независимых операций между различными машинами. Однако, следует помнить, что далеко не все существующие алгоритмы могут быть эффективно разделены на независимые друг от друга операции, что может накладывать некоторые ограничения на её повсеместное использование.

Список использованных источников:

1. Dean, J., Ghemawat, S., 2004, *MapReduce: Simplified Data Processing on Large Clusters* [Электронный ресурс]. – Электронные данные. – Режим доступа: <http://research.google.com/archive/mapreduce-osdi04.pdf>
2. *ApacheHadoop* [Электронный ресурс]. – Электронные данные. – Режим доступа: <http://hadoop.apache.org>
3. *HDFS Architecture Guide* [Электронный ресурс]. – Электронные данные. – Режим доступа: [http://hadoop.apache.org/docs/stable1/hdfs\\_design.html](http://hadoop.apache.org/docs/stable1/hdfs_design.html)

## **ВЫБОР И ПРИМЕНЕНИЕ МЕТОДОВ МОНЕТИЗАЦИИ ДЛЯ ИНФОРМАЦИОННО-СПРАВОЧНЫХ ИНТЕРНЕТ-ПРОЕКТОВ**

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Павлюченко Д. О.*

*Пешков А. Т. – канд. техн. наук, доцент*

В настоящее время методы монетизации и их применение в интернет-проектах это важный аспект существования любого сайта. От этого зависит будущее проекта и его перспективы на рынке, так как монетизация является основным источником прибыли, получаемой от интернет-проектов.

В зависимости от вида интернет-проекта, подразумевается использование определенных методов монетизации, которые наилучшим образом подходят и используют весь потенциал данного интернет-проекта.

Информационно-справочные интернет-проекты представляют собой сайты в сети интернет основной задачей которых является предоставление информации пользователям в виде каталогов услуг/предприятий, сборников тематических материалов, сборников медиа-информации. Такие сайты могут быть общей тематики (например, каталог товаров и услуг города Минска) или узконаправленной тематики (например, каталог аудиторских и бухгалтерских компаний).

Основной метод монетизации, применимый к данному виду сайтов - это монетизация предоставляемой информации, как для потребителей, так и для поставщиков этой информации.

Учитывая особенностей информационно-справочных интернет-проектов для анализа был выбран ряд методов монетизации, таких как монетизация информационно-справочного каталога (основная информация сайта), монетизация посредством размещения статей и контекстных ссылок, монетизация с помощью баннерной рекламы, монетизация с помощью партнерских программ.

Основные выводы по основе результатов анализа применения выбранных методов монетизации:

1. Монетизация информационно-справочного контента сайта является основным источником дохода для данного вида интернет-проектов - 80% прибыли получено при применении этого метода.