УДК 004.67

А.А.Перескокова, А.А.Кольцов, В.И.Кочугова, С.Толпыгин, И.В.Никифоров,

АРХИТЕКТУРА ПРОГРАММНОГО АНАЛИТИЧЕСКОГО КОМПЛЕКСА ИССЛЕДОВАНИЯ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ НА ОБРАЗОВАТЕЛЬНОЙ ПЛАТФОРМЕ “ОТКРЫТОЕ ОБРАЗОВАНИЕ”

По всему миру существует множество образовательных платформ, распространяющих электронные курсы. Наиболее популярные из них в РФ:

* «Открытое образование» - современная образовательная платформа, предлагающая онлайн-курсы по базовым дисциплинам,   
  изучаемым в российских университетах[1];
* Moodle - истема управления курсами, также известная как система управления обучением или виртуальная обучающая среда[2].

Все образовательные платформы имеют возможность сохранять активность (действия) пользователей на сайте во время прохождения обучения на курсе или выполнения тестовых и экзаменационных заданий (ведение аудита действий пользователя). Но далеко не все образовательные платформы предоставляют возможность анализировать активность пользователей и преподавателей с целью совершенствования контента курса, ресурса и платформы.

Платформа «Открытое образование» - одна из наиболее активно использующихся систем образования всей России, в том числе СПбПУ, и, к сожалению, имеет крайне ограниченный набор аналитических возможностей.

Поэтому актуальной является задача разработки программного аналитического комплекса для платформы «Открытое Образование».

Платформа позволяет:

* Логгировать каждое действие пользователя (студента, преподавателя, администратора) в формате JSON.
* Вручную скачать аудит в виде лог-файла, однако, необходим специальный доступ. Не каждому пользователю доступно такое действие.

Сопутствующие ограничения:

* Большой объем файлов лога (например, 18Гб, 100 000 000 записей)
* невозможность человеку вручную проанализировать лог файлы и собрать статистику
* открыть файл для просмотра стандартными программными средствами
* Наличие «пустых», неинформативных строк
* Наличие более 200 разных типов событий пользователей

Необходимо предоставить возможность анализа логов за счет создания программного средства, объединяющего в себе выгрузку, хранение и обработку.

Целью данной работы является предоставление возможности проведения анализа логов портала «Открытое образование» за счет создания программного аналитического средства, реализующего в себе выгрузку, хранение, обработку и отображение результатов анализа.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести исследование и сравнить технологии анализа логов.
2. Исследовать структуру и формат логов, предоставляемых образовательной платформой «Открытое образование».
3. Предложить программную структуру программного средства, позволяющего анализировать логи.
4. Реализовать модули парсинга и анализа логов.
5. Продемонстрировать работу приложения.

Логи представляют собой файл, в каждой строчке которого содержится отдельный JSON-объект, для которого: каждый пользователь имеет уникальный идентификатор, каждая строчка лога представляет собой отдельное событие. При этом каждое событие имеет свой тип, типы строго задокументированы платформой, например:

edx.course.enrollment.activated - событие, при котором студент записался на курс;

edx.course.enrollment.deactivated - событие, при котором студент вышел из курса.

А также, вложенные структуры различны в зависимости от события - обрабатывая и комбинируя анализ различных событий можно получать «интересные» результаты.

На начальном этапе была предложена архитектура приложения (рис.1), при которой все вычислительные компоненты находятся в одном и том же месте и предоставляют пользователю простой интерфейс. При этом инструмент анализа скачивает необходимые логи, обращаясь к образовательной платформе, которая находится на удаленном сервере.

Система может быть установлена локально – для демострации работоспособности, и на удаленном сервере с мощными аппаратными возможностями для достижения наилучшей производительности. Для установки системы локально потребуется операционная система Windows.

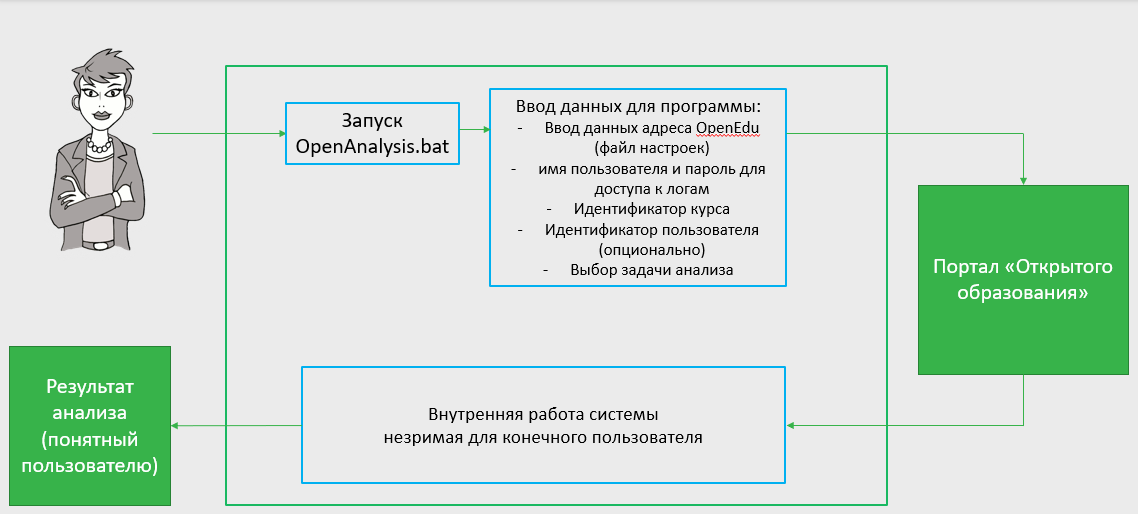


Рисунок 1. Укрупненная архитектура инструмента анализа логов.

На рис.2 представлена внутренняя архитектура инструмента анализа логов. Инструмент включает в себя несколько независимых модулей:

* Модуль аутентификации пользователя на портале OpenEdu
* Модуль выгрузки логов курса в .zip архив;
* Модуль предобработки (фильтрация и парсинг) логов;
* Модуль записи логов в аналитическую базу данных;
* База данных PostgreSQL[3];
* Модуль анализа логов, включающий некоторое множество задач;
* Пользовательский интерфейс, предоставляющий возможность выбора задачи анализа и отображающий(сохраняющий) результаты.

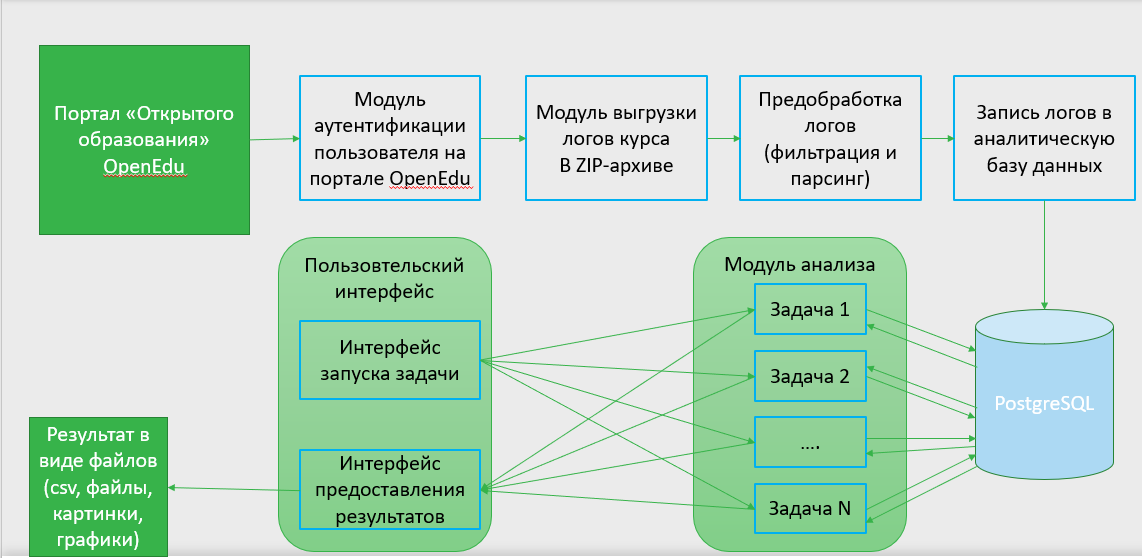


Рисунок 2. Внутренняя архитектура инструмента анализа логов.

Данная архитектура позволила создать инструмент анализа логов платформы «Открытое образование» для получения метрик по образовательным курсам. С помощью разработанного инструмента можно выгружать лог прямо с образовательной платформы, фильтровать его от неинформативных данных, уменьшая его объем и облегчая анализ, запускать различные задачи анализа, предоставлять и сохранять результаты анализа.

Для тестирования работоспособности прототипа системы был получен обфускированный лок прохождения обучения 4000 пользователей по одному из доступных курсов. Изначальный размер архивированного лог-файла 1.5Гб, после распаковки JSON представление занимает 18Гб и содержит около 18 000 000 записей. Загрузка записей в БД PostgreSQL занимает около 60 минут, а за счет использования индексации и орагнизации данных по таблицам, исполнение аналитических запросов осуществляется достаточно быстро (до 60 секунд).

В текущий момент система находится в виде прототипа и список выполняемых аналитических задач планируется занчительно расширять.

ЛИТЕРАТУРА

1. Открытое образование [Электронный ресурс]: <https://openedu.ru/>
2. Образовательная среда Moodle [Электронный ресурс]: <https://moodle.org/>
3. База данных PostgreSQL [Электронный ресурс]: <https://www.postgresql.org/>