

Программа log_analyzer

(Домашнее задание OTUS Python, студент Сергей Симонов)

1 Назначение программы

1.1 Программа предназначена для анализа файлов протоколирования (далее, лог-файлов), сгенерированных модулем [ngx_http_log_module](#) с целью выявления проблемных ресурсов web-сервиса.

1.2 Программа анализирует последний лог-файл и генерирует отчет в формате html.

2 Входные данные и параметры конфигурации

2.1 Входными данными являются лог-файл и файл конфигурации.

2.2 Программа разработана и протестирована для следующего формата протоколирования:

```
$remote_addr    $remote_user  $http_x_real_ip  [$time_local] "$request" '
$status         $body_bytes_sent    "$http_referer"     "$http_user_agent"
"$http_x_forwarded_for"    "$http_X_REQUEST_ID"    "$http_X_RB_USER"
$request_time
```

Формат протоколирования может быть изменен, однако необходимо, чтобы присутствовали поля "\$request" и далее \$request_time.

2.3 Лог-файл должен иметь кодировку UTF-8. Допускается помещать файлы в архив gzip. Имена лог-файлов должны соответствовать формату:

nginx-access-ui.log-YYYYMMDD или nginx-access-ui.log-YYYYMMDD.gz (для архива)

2.4 В параметре командной строки программе указывается путь к файлу конфигурации. Если файл не указан, то используется файл `./log_analyzer.cfg`.

2.5 Файл конфигурации представляет собой текстовый файл в кодировке (UTF-8) в котором указаны параметры в формате JSON. Назначение параметров приведено в таблице 1.

2.6 Если в файле конфигурации не задан какой-либо параметр, то используется значение по умолчанию.

Таблица 1. Параметры работы программы

Параметр	Значение по умолчанию	Описание
REPORT_SIZE	1000	Количество URL с наибольшим суммарным временем обработки в отчете.
REPORT_DIR	«./reports»	Путь к папке, в которой помещаются отчеты (результат анализа). В папке должен находиться файл report.html (шаблон) и файлы jquery.tablesorter.js, jquery.tablesorter.min.js
LOG_DIR	«./log»	Путь к папке, в которой находятся лог-файлы
LOGGER_FILE NAME	None	Путь к файлу с отчетом работы программы. Если не указан, то осуществляется вывод на экран

3 Выходные данные

3.1 Для последнего лог файла, создается отчет в формате `report-YYYYMMDD.html`, который помещается в папку REPORT_DIR (указывается в файле конфигурации).

3.2 При просмотре отчета в браузере таблица строки таблицы могут быть отсортированы. Для этого необходимо нажать на заглавие столбца.

3.3 Ход работы и ошибки выполнения программы выводятся на экран или, если задан `LOGGER_FILENAME`, в соответствующем файле отчета работы программы.

4 Запуск программы

4.1 Программа представляет собой скрипт для Python 3.X. Для запуска должен быть установлен соответствующий интерпретатор.

4.2 Для запуска программы выполнить команду

```
python log_analyzer.py
```

в этом случае используется файл конфигурации по умолчанию, или

```
python log_analyzer.py config='путь к файлу конфигурации'
```

5 Запуск тестов

5.1 Для программы разработана система тестов на базе unittest (файл `log_analyzer_tests.py`). Необходимые для тестов файлы размещены в папке tests.

5.2 Для запуска тестов используйте скрипт `log_analyzer_runtests.py`:

```
python log_analyzer_runtests.py
```