**Программа log\_analyzer**

*(Домашнее задание OTUS Python, студент Сергей Симонов)*

1. Назначение программы
   1. Программа предназначена для анализа файлов протоколирования (далее, лог-файлов), сгенерированных модулем [ngx\_http\_log\_module](http://nginx.org/en/docs/http/ngx_http_log_module.html) с целью выявления проблемных ресурсов web-сервиса.
   2. Программа анализирует последний лог-файл и генерирует отчет в формате html.
2. Входные данные и параметры конфигурации
   1. Входными данными являются лог-файл и файл конфигурации.
   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*.*

* 1. 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)

* 1. В параметре командной строки программе указывается путь к файлу конфигурации. Если файл не указан, то используется файл ./log\_analyzer.cfg.
  2. 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 находится 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\_FILENAME | None | Путь к файлу с отчетом работы программы. Если не указан, то осуществляется вывод на экран |

1. Выходные данные
   1. Для последнего лог файла, создается отчет в формате   
      report-YYYYMMDD.html, который помещается в папку REPORT\_DIR (указывается в файле конфигурации).
   2. При просмотре отчета в браузере таблица строки таблицы могут быть отсортированы. Для этого необходимо нажать на заглавие столбца.
   3. Ход работы и ошибки выполнения программы выводятся на экран или, если задан LOGGER\_FILENAME, в соответствующем файле отчета работы программы.
2. Запуск программы
   1. Программа представляет собой скрипт для Python 3.X. Для запуска должен быть установлен соответствующий интерпретатор.
   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’

1. Запуск тестов
   1. Для программы разработана система тестов на базе unittest (файл log\_analyzer\_tests.py). Необходимые для тестов файлы размещены в папке tests.
   2. Для запуска тестов используйте скрипт log\_analyzer\_runtests.py:

python log\_analyzer\_runtests.py