

README.md

Задание

MemcLoad

Задание: нужно переделать однопоточную версию memc_load.py в более производительный вариант. Сам скрипт парсит и заливает в мемкеш поминутную выгрузку логов трекера установленных приложений. Ключом является тип и идентификатор устройства через двоеточие, значением является protobuf сообщение.

```
$ ls -lh /Users/s.stupnikov/Coding/otus/*.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:09 /Users/s.stupnikov/Coding/otus/20170929000000.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:17 /Users/s.stupnikov/Coding/otus/20170929000001.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:25 /Users/s.stupnikov/Coding/otus/20170929000002.tsv.gz

$ gunzip -c /Users/s.stupnikov/Coding/otus/20170929000000.tsv.gz | head -3
idfa      e7e1a50c0ec2747ca56cd9e1558c0d7c      67.7835424444      -22.804400547
idfa      f5ae5fe6122bb20d08ff2c2ec43fb4c4      -104.68583244      -51.24448376
gaid      3261cf44cbe6a00839c574336fdf49f6      137.790839567      56.8403675248
```

Ссылки на tsv.gz файлы:

- <https://cloud.mail.ru/public/2hZL/Ko9s8R9TA>
- <https://cloud.mail.ru/public/DzSX/oj8RxGX1A>
- <https://cloud.mail.ru/public/LoDo/SfsPEzoGc>

Пример запуска

```
$ python memc_load.py --pattern=/Users/s.stupnikov/Coding/otus/*.tsv.gz --dry
[2017.09.29 12:39:52] I Memc loader started with options: {'dry': True, 'log'
[2017.09.29 12:39:52] I Processing /Users/s.stupnikov/Coding/otus/20170929000
[2017.09.29 12:39:52] D 127.0.0.1:33013 - idfa:e7e1a50c0ec2747ca56cd9e1558c0d
[2017.09.29 12:39:52] D 127.0.0.1:33013 - idfa:f5ae5fe6122bb20d08ff2c2ec43fb4
[2017.09.29 12:39:52] D 127.0.0.1:33014 - gaid:3261cf44cbe6a00839c574336fdf49
```

Важно обрабатывать файлы в хронологическом порядке. В данном случае под этим имеется в виду, что после обработки нужно переименовывать файл, префиксировав имя точкой, последовательно и хронологически. Заливать же можно параллельно.

Цель задания: реализовать типичную задачу конкурентной обработки данных, поэкспериментировать с различными подходами, получить опыт взаимодействия со стандартной библиотекой.

Критерии успеха: задание **обязательно**, критерием успеха является работающий согласно заданию код, для которого написаны минимальные тесты, проверено соответствие пер8, написана документация с примерами запуска (боевого и тестов), в README, например. Далее успешность определяется code review.

Deadline

Задание нужно сдать через неделю. То есть ДЗ, выданное в понедельник, нужно сдать до следующего занятия в понедельник. Код, отправленный на ревью в это время, рассматривается в первом приоритете. Нарушение дедлайна (пока) не карается, пытаться сдать ДЗ можно до конца курса. Но код, отправленный с опозданием, когда по плану предполагается работа над более актуальным ДЗ, будет рассматриваться в более низком приоритете без гарантий по высокой скорости проверки

Обратная связь

Студент коммитит все необходимое в свой github/gitlab репозиторий. Далее необходимо зайти в ЛК, найти занятие, ДЗ по которому выполнялось, нажать "Чат с преподавателем" и отправить ссылку. После этого ревью и общение на тему ДЗ будет происходить в рамках этого чата.