Техническое задание на разработку инструмента

краулер отзывов Google Maps

Москва, 26 июня 2019 г.

1. Требования к функциональности

1.1. Задача и вход/выход программы

Краулер решает задачу сбора данных, оценок и отзывов об организациях, находящихся в Google Maps по определенному запросу.

1. **Входные данные**: строка длины, не превышающей *n* в кодировке utf-8, где *n* задается в 1.3;

2. **Выходные данные**: информация о найденных по входному запросу компаниях с информацией о ее названии, местоположении, пользовательской оценке;

3. **Формат и запись выходных данных**: модуль записи выходных данных должен быть абстрагирован и поддаваться легкой модификации для записи в различные базы данных. В текущей реализации инструмент записывает данные в .csv файл с оговоренными полями.

4. **Информация о компании (поля)**: id, название, местоположение, подзаголовок с описанием рода деятельности компании, пользовательская оценка (звезды);

5. **Пользовательская оценка (поля)**: reviewId, информация об отзыве;

6. **Информация об отзыве**: автор, кол-во отзывов автора на момент сбора, время сбора, дата публикации, рейтинг, заголовок, текст, количество лайков. Текст отзыва должен быть минимально предобработан (очищен от html-сущностей, фрагментов кода и т.п.). Текст должен быть приведен полностью, либо, если текст превышает указанную максимальную длину текста *l*, приводиться до последнего предшествующего *l* пробела.

1.2. Запуск и время работы

1. Средняя скорость обработки одного запроса от ввода ключевых слов до записи собранных и предобработанных данных в файл в среднем не превышает *m* минут, где *m* задается в 1.3.

2. Поддерживается возможность одновременного запуска нескольких экземпляров краулера.

3. Поддерживается возможность многопоточного сбора по одному запросу (количество потоков – настраиваемый пользовательский параметр).

1.3. Ограничения работы

n	100	*длина запроса, символы*
m	5	*средняя скорость работы, мин*
l	8000	*максимальная длина текста отзыва*

2. Требования к реализации

1. Реализация не должна нарушать условия работы с Google Maps;
2. Краулер должен быть передан заказчику в виде исходного кода с необходимыми комментариями.