

# Техническое задание на разработку инструмента

«Краулер отзывов из картографических

источников»: Google Maps, TripAdvisor, 2GIS, Yandex.maps

## 1. Требования к функциональности

Инструмент решает задачу сбора и парсинга данных отзывов из

источников:

- Google Maps
- 2GIS
- TripAdvisor
- Yandex.maps

### 1.1. Информационные поля

Информация из источников собирается единообразно

Информация об организации в определенной точке

1. Местоположение (координаты) организации записывается в БД в поле orgLoc в едином формате.
2. Текстовый адрес организации записывается в поле orgAddress, в кириллице.
3. Название организации записывается в поле orgName, в кириллице.
4. Рейтинг организации в данной точке записывается в поле orgStars в виде числового значения с одним знаком после запятой.
5. Информация о категории организации(ее сфере деятельности) в таксономии источника записывается в поле orgType. В случае, если источник присваивает организации более одного типа, они записываются в поле через разделитель.
6. Информация о сайте организации, при ее наличии, записывается в поле orgSite.
7. Вся информация о дополнительных признаках организации (оплата, предложения, особенности заведения и т.п.) записывается в поле orgFeatures в формате feature1: value1, value2; feature2: value3 и т.д.
8. Источник записывается в поле source в виде ссылки на организацию.

## **Информация об отзыве**

1. Каждому отзыву присваивается уникальный идентификатор и записывается в поле `reviewid`.
2. Текст отзыва предобрабатывается и записывается в поле настраиваемой длины `reviewText`
3. Дата и время сбора отзыва соответствует моменту отправки батча отзывов на запись в файл/БД и сохраняется в поле `reviewDTTM`
4. Дата публикации отзыва записывается в поле `reviewPubDate` в едином формате `DD.MM.YYYY`.
5. Рейтинг отзыва соответствует количеству звезд и записывается в поле `reviewStars`.

## **2. Требования к реализации**

1. Модуль записи собранных данных абстрагирован от логики сбора и предобработки и модифицируется под запись в различные БД;
2. Краулер поддерживает многопоточный сбор и запись;
3. Можно запустить несколько копий краулера одновременно;
4. Краулер сдается заказчику в виде исходного кода на Python3;
5. Сбор 1000 отзывов осуществляется не более, чем за 60 сек.