**по курсу:**

**«Обработка естественно-языковых текстов»**

**по теме:**

**«Построение сниппетов»**

**2 семестр**

Студент: Ахмед С. Х.

Преподаватель: Калинин А. Л.

Группа: 8О-106М

**Москва, 2019 г**

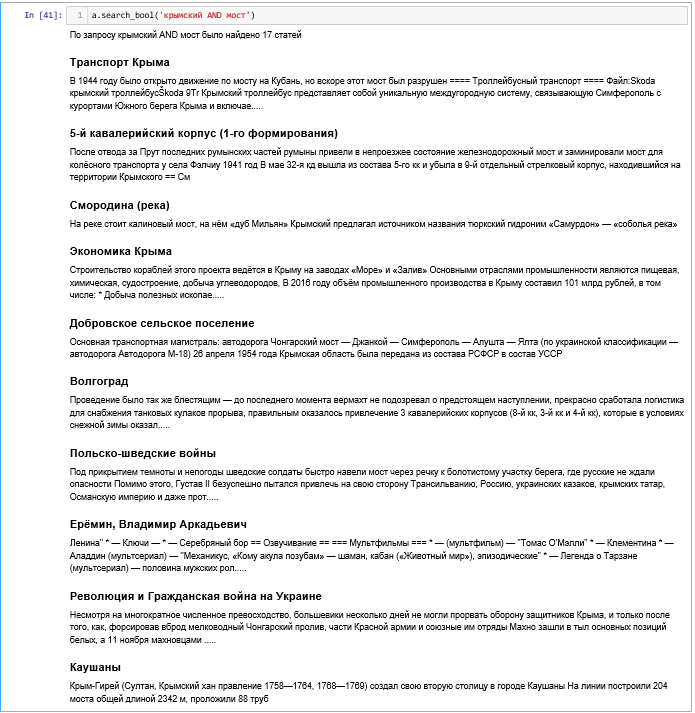
Постановка задачи

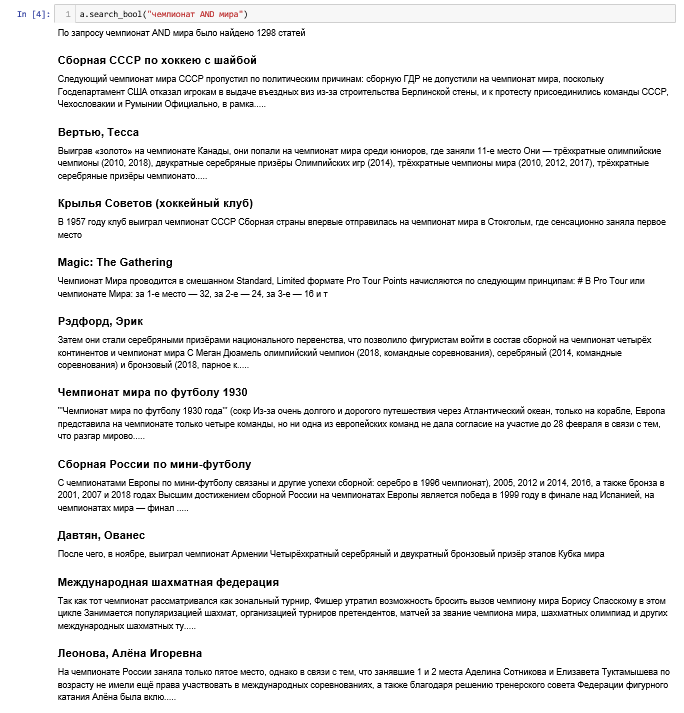
Необходимо добавить в поисковую систему построение цитат (сниппетов), реферирование документов, найденных по запросу. Сниппеты должны содержать слова запроса и давать пользователю представление о том, насколько документ отвечает поисковому запросу. Длина сниппета должна быть ограничена двумя-тремя строчками. В отчёте нужно привести описание алгоритма построения сниппетов, примеры.

Методика решения

Для построения сниппетов информации, которая хранится в индексе недостаточно, необходима доп. Информация, для предотвращения дополнительных действий. Я решил создать структуру при индексации, которая хранит в себе информацию о номере первого предложения, в которое входит термин из запроса. Сам алгоритм прост. Итерируясь по терминам запроса и результирующим документам собираются предложения, потом сортируются по позициям, если необходимо итоговое предложение урезается по длине(300 символов), а затем выдаются пользователю.

Примеры





Выводы

Выявлены следующие недочеты:

1. Длина текста в некоторых моментах мала
2. Ошибки форматирования текста при парсинге(остались некоторые служебные символы)
3. Не учитывается момент связанный с tf оценкой термина.

Построенные сниппеты отображают содержание статьи, однако не учитывают значимость термина для запроса и документа