



Лабораторная работа №1 NLP для интеллектуального поиска структурированной информации в «сырых» данных

Необходимо из выходных данных публикаций, собранных с сайта Scopus в формате json, в котором присутствуют поля

'name' – название публикации;
'authorlist' – список авторов;
'abstract' – текст аннотации статьи;
'journalInfo' – выходные данные статьи, включая название издания, номер тома, с
 страницы, год издания;
'affiliationlist' – список аффилиаций авторов статьи;
'authorlistLinks' – список ссылок на профили авторов в Scopus (включает в себя
 Scopus ID каждого автора).

сформировать и заполнить таблицу в формате csv, в которой были бы следующие столбцы: ['Имя автора', 'Scopus ID', 'Год', 'Количество публикаций']. В файле **Example_lab1.ipynb** содержится пример поиска имени автора в полях 'authorlist' в файле **user_articles_data.json** для формирования списка всех доступных аннотаций статей этого автора. Необходимо выполнить поиск всех статей каждого автора, рассортировать их по годам издания и заполнить указанную таблицу.

Лабораторная работа №2 Методы машинного обучения для оценки схожести научных интересов

В файле **Example_lab2.ipynb** приведен метод расчета представлений пользователей на основе аннотаций их статей из файла **arts.json**; затем на основе этих представлений рассчитывается научная совместимость авторов. Вам предлагается определить совместимость на основе графа соавторств с помощью метода `node2vec` (`node2vec: Scalable Feature Learning for Networks`. A. Grover, J. Leskovec. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016.). В файле **auth.emb** находятся представления авторов, рассчитанные с помощью `node2vec`. Используйте их для того, чтобы рассчитать похожесть. Сравните результаты. Какой метод для определения научной совместимости можете предложить вы?