

Домашнее задание 3

Темы

Задание можно выполнять в группах, можно по одному. В [таблице](#) есть список тем и описания заданий, запишитесь, пожалуйста, в последней колонке, кто какую тему берет.

В колонке "количество человек в группе" указано, на сколько человек рассчитана тема. В колонке "Мак. количество групп" указано, сколько групп могут взять это задание. Последнее задание рассчитано на одного человека и его может взять сколько угодно человек.

Требования к выполнению задания

- 1) Не забудьте про предобработку, в списке материалов есть ссылки на списки стоп-слов, скрипт для лемматизации, а также файлы со списками лемм.
- 2) Работу нужно оформить в виде презентации на 5-7 слайдов (для групп) или 3-5 слайдов (для тех, кто работает один). Опишите ваши данные, ход работы (какие стоп-списки использовали, как лемматизировали, как строили списки и т.п.), полученные результаты и наблюдения.
- 3) Помимо презентации пришлите мне в любом удобном для вас формате списки, которые у вас получились. В AntConc, например, в *File* есть опция *Save output*. Кто будет работать с Python могут прислать код с пометками, где какой список.

Материалы

- 1) Для заданий по RAKE алгоритму <https://pypi.org/project/rake-nltk/>
- 2) Для задания по TextRank <https://github.com/summanlp/textrank>
- 3) Программа AntConc www.laurenceanthony.net/software/antconc/
- 4) Как работать в AntConc, в том числе описание вкладок Worlists и Keywords <https://wmtang.org/corpus-linguistics/antconc-tutorial/>
- 5) Файлы со стоп-словами (en_ru_stopwords.zip) <https://github.com/NeLyu/djscompling/tree/master/assignment3>
- 6) Python стоп-слова <https://pypi.org/project/stop-words/>
- 7) Список лемм для AntConc http://www.laurenceanthony.net/resources/wordlists/antbnc_lemmas_ver_003.zip
- 8) BNC корпус, частотные списки http://www.laurenceanthony.net/resources/wordlists/bnc_wordlist.zip
- 9) Лемматизация для английского и русского (Lemmatization_en_ru.ipynb) <https://github.com/NeLyu/djscompling/tree/master/assignment3>
- 10) Тексты на английском по тематикам (Texts_by_topic.zip) <https://github.com/NeLyu/djscompling/tree/master/assignment3>