

noSketchEngine

Что позволяет делать noSketchEngine со специально подготовленным корпусом текстов?

1. Конкорданс (поиск по словоформе, лемме, фразе или с помощью специального языка CQL -- corpus query language¹)
2. Фильтрация конкорданса
Фильтрация используется для того, чтобы уточнить (сузить) результаты выдачи. Можно указывать как обязательные условия, по которым будет производиться отбор (позитивная фильтрация), так и невозможные условия – соответствующие им явления не будут включены в итоговую выдачу (негативная фильтрация)².
3. Сортировка конкорданса (по левому/правому контексту, по метаданным документов)
Под сортировкой понимается возможность выбора порядка представления результатов поисковой выдачи по различным параметрам (отличается от фильтрации тем, что не предполагает изменения количества примеров, а лишь меняется порядок их представления).
4. Подсчет и представление частотных характеристик (ipm, ARF и др.) заданной единицы по различным параметрам:
 - Одноуровневое распределение - распределение словоформ, лемм исходного запроса; распределение заданной единицы по документам, авторам, по полу говорящего, по типам текста, жанрам, тематике и другим метаданным, выбираемых пользователем.
 - Многоуровневое распределение - распределения заданной леммы и единиц, составляющих её правый и/или левый контекст.
5. Коллокации
Коллокации - это две и более лексические единицы, которые характеризуются высокой частотой совместной встречаемости. В качестве атрибута для поиска коллоката к заданному слову или фразе можно выбрать слово, лемму, частеречный тег и некоторые другие параметры. В качестве статистических показателей используются такие меры, как T-Score, MI-Score, MI3-Score, log-likelihood, log-dice, salience, minimum sensitivity, relative frequency.

¹ См. подробнее по ссылке <https://www.sketchengine.co.uk/corpus-querying/>

² См. подробнее о фильтрации, сортировке, коллокациях и о подсчете частотных характеристик по ссылке

<https://www.sketchengine.co.uk/user-guide/user-manual/concordance-introduction/working-with-concordance-results/>

6. Частотные списки слов

Данная опция предполагает построение списка лемм/словоформ, отсортированных по частоте. Частота (ipm, ARF, документная) считается по задаваемым пользователем параметрам (подкорпус, взятый за основу, black and white lists, минимальные и максимальные частоты слов и др.)³

7. Возможность сохранения собственных подкорпусов.

8. Статистика по корпусу.

³ См. подробнее по ссылке <https://www.sketchengine.co.uk/user-guide/user-manual/word-list/>