8 Работа с параллельными корпусами

Задание 1 (1/2/3 уровень)

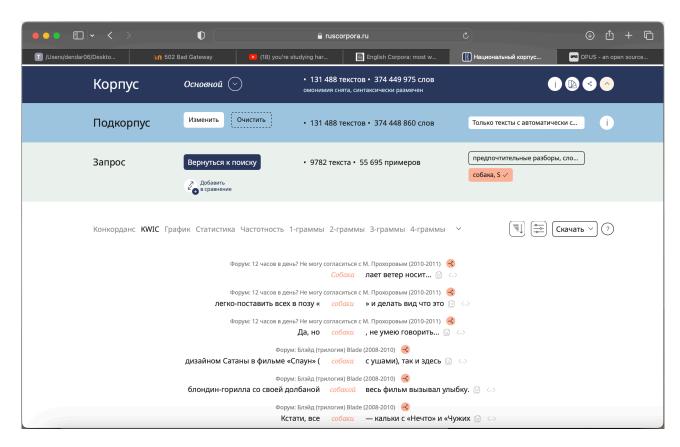
Аналогичные ресурсы:

- 1. **The Leipzig Interlinear Corpus** (https://corpora.informatik.uni-leipzig.de/interlinear/) корпус, содержащий тексты на многих языках с переводом на английский язык.
- 2. **The Open Parallel Corpus** (https://opus.nlpl.eu/) большой параллельный корпус, содержащий миллионы параллельных предложений на более чем 100 языках.
- 3. **BabelCorpus** (https://babelcorp.io/) бесплатный онлайн-корпус, который содержит параллельные тексты на более чем 50 языках.
- 4. **The United Nations Parallel Corpus** (http://www.un.org/en/universal-declaration-human-rights/ungmcorpus.html) параллельный корпус, который содержит тексты Организации Объединенных Наций на более чем 6000 языках.

Анализ:

• **НКРЯ** - это крупнейший корпус русского языка, содержащий более 150 миллионов словоупотреблений.

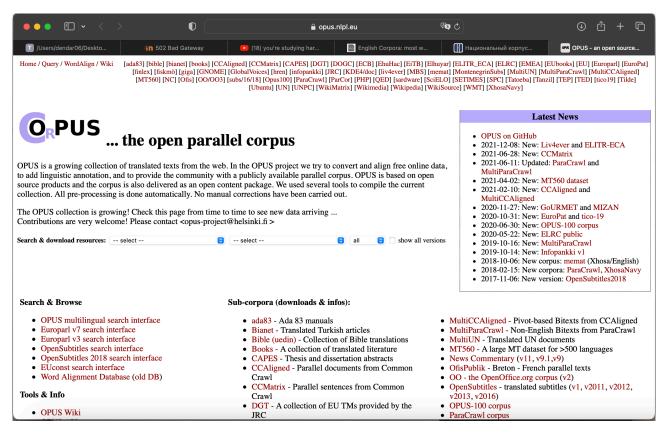
Для изучения лингвоспецифической лексики в НКРЯ можно использовать поиск по ключевым словам или тематические разделы, чтобы найти тексты, содержащие специализированную лексику. Например, если вас интересует медицинская терминология, можно искать соответствующие статьи в разделе научной литературы. После этого можно анализировать контекст, в котором употребляются специальные термины, и углубляться в изучение соответствующей лексики.



Для этого можно использовать различные инструменты анализа текстов, предоставляемые НКРЯ, такие как частотность употребления слов, конкордансы и морфологический анализ. Эти инструменты могут помочь выявить особенности употребления специализированной лексики в контексте.

• OPUS - это международный корпус параллельных текстов, содержащий более 200 миллионов словоупотреблений на более чем 20 языках. Это параллельный международный корпус, который содержит тексты на нескольких языках. OPUS позволяет исследователям сравнивать лингвоспецифическую лексику разных языков и изучать их сходства и различия.

OPUS также предоставляет различные инструменты для анализа текстов, такие как частотность употребления слов, выравнивание параллельных текстов и другие, что делает его полезным ресурсом для изучения лингвоспецифической лексики в различных языковых контекстах.



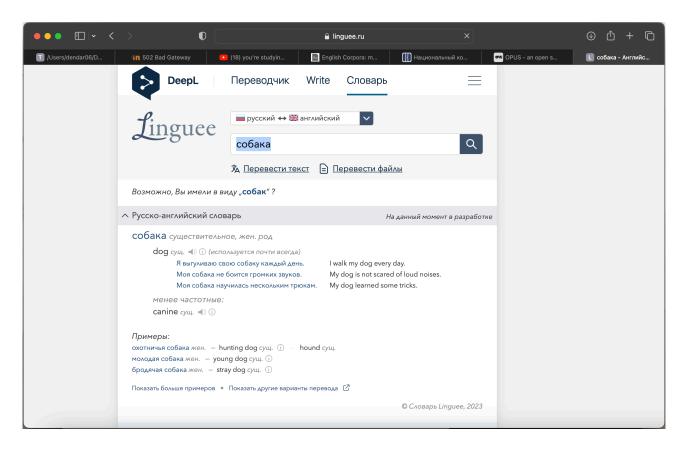
Linguee - это онлайн-словарь, который содержит примеры использования слов и фраз на различных языках. Linguee позволяет исследователям находить лингвоспецифические выражения и изучать их контекст использования.

Функционал и возможности Linguee:

- 1. Поиск по ключевым словам: Введите термин, который вам интересен, и получите примеры его употребления в параллельных текстах на разных языках. Это поможет понять, как термин переводится и как используется в различных контекстах.
- 2. Изучение специализированных терминов: Linguee может быть полезен для изучения специальных терминов в различных областях, таких как медицина, юриспруденция, техника

и другие. Просматривайте примеры употребления терминов в различных контекстах, чтобы получить представление о их специфике.

3. Сравнение контекстов использования: Linguee позволяет сравнивать, как определенное слово или выражение используется в различных языках. Это помогает понять, какие нюансы имеются в употреблении терминов в разных языках и как они соотносятся между собой.



Задание 2 (2/3 уровни)

Программа	Особенности программы
LEOBILINGUA 1.0	Интуитивно понятный интерфейс, автоматическое выравнивание текста, поддержка нескольких языков, встроенный словарь для перевода и анализа выравнивания.
	Минусы: - Возможные ошибки в автоматическом выравнивании Ограниченные возможности: Возможно, программа не предоставляет некоторых продвинутых функций, которые пользователь может потребовать для конкретной задачи Необходимость постоянного обновления: Для поддержания актуальности и своевременного исправления возможных ошибок требуется
Hunalign	Плюсы программы:
	- Автоматическое выравнивание
	последовательностей
	- Высокая точность выравнивания - Возможность работы с большими
	объемами данных
	- Простой и понятный интерфейс
	Минусы программы:
	- Программа может быть медленной на
	больших данных
	- Требуется достаточное количество памяти
	для работы программы
	- Не все типы последовательностей
	поддерживаются программой
Champollion	Особенности программы:
	- Программа работает только с текстовыми файлами.
	- Champollion поддерживает только английский язык.
	- Программа требует определенного
	формата входного текста.
	- Время работы программы может быть
	достаточно большим, особенно при работе
	с большими текстами.

Skuuper Cleaner	Особенности программы: - Программа может удалять не только шум и ошибки, но и полезную информацию. Поэтому перед очисткой текста рекомендуется сохранить его копию Настройки очистки можно изменять, чтобы получить более качественный результат. Однако слишком сильное изменение параметров может привести к удалению полезной информации Программа работает достаточно быстро, но на больших текстах может потребоваться некоторое время.
YouAlign	Плюсы программы: - Возможность быстро найти похожие тексты в интернете - Простота использования - достаточно ввести текст и нажать кнопку - Высокая точность результатов Минусы программы: - Требуется подключение к интернету - Может не работать с некоторыми типами текстов - Может выдавать нежелательные результаты