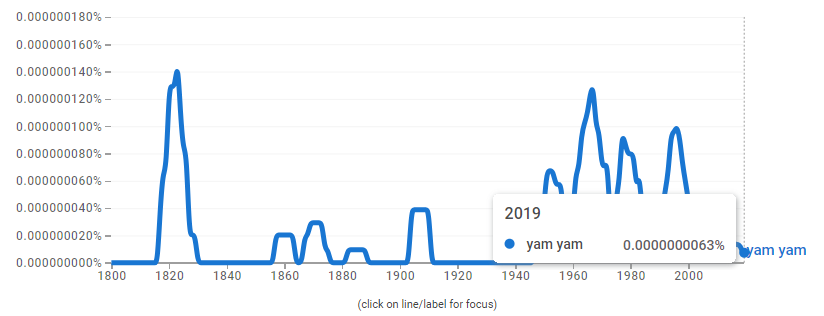
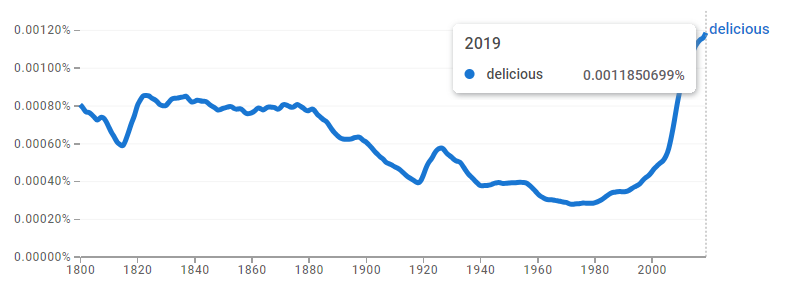
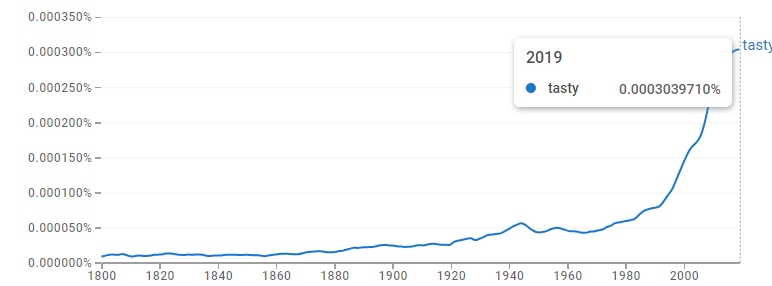
**Уровень 1**

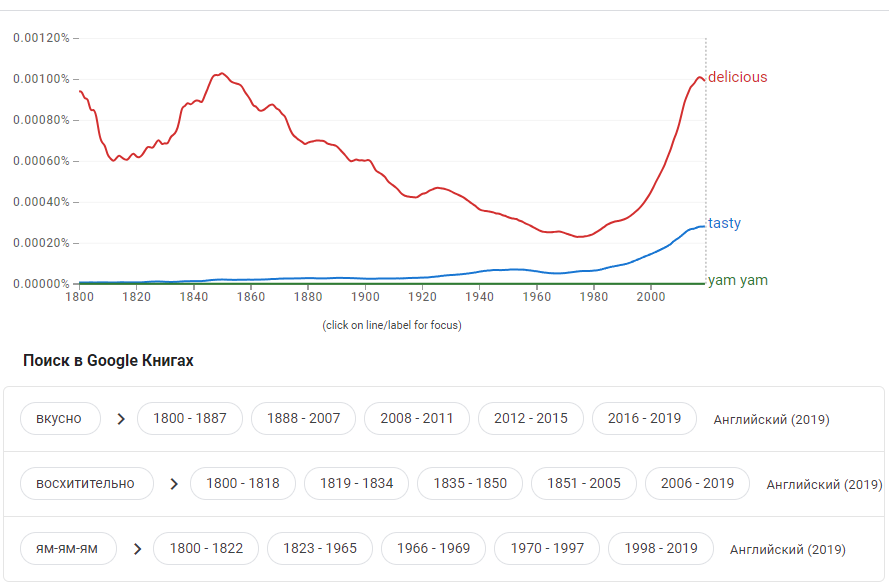
**1.**

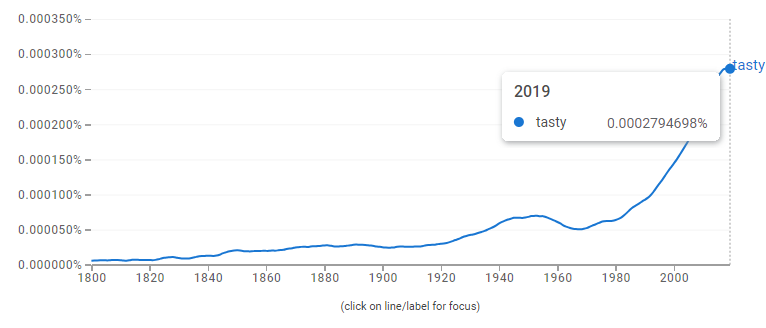
Британский

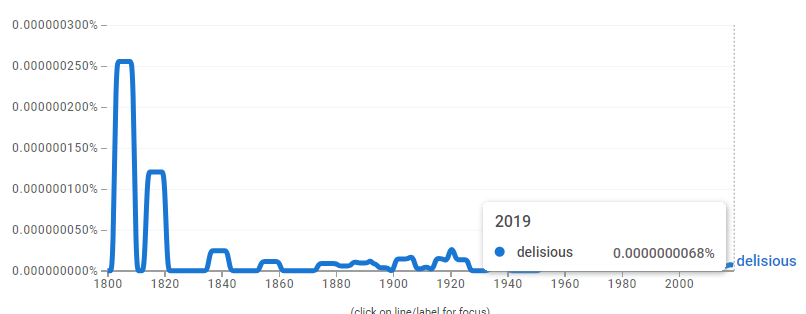


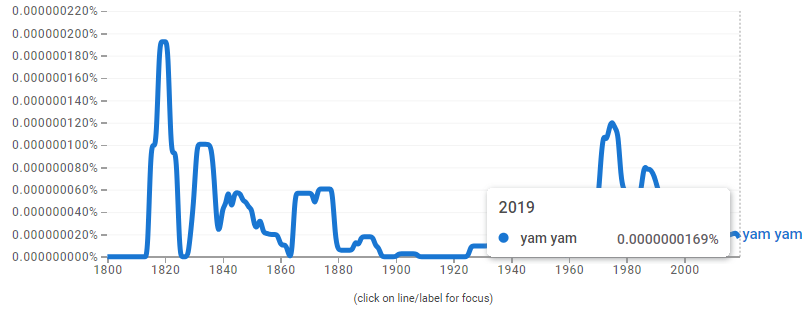


Американский

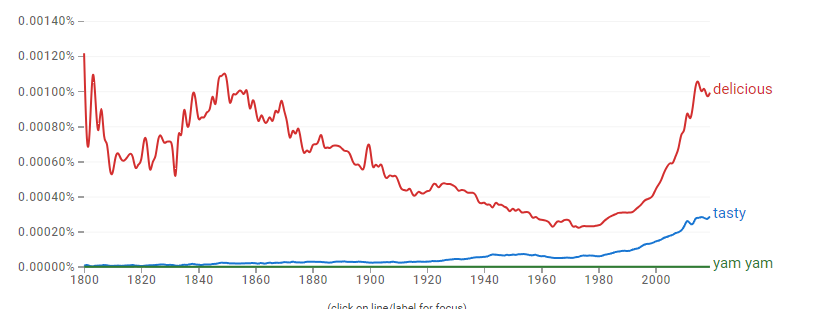


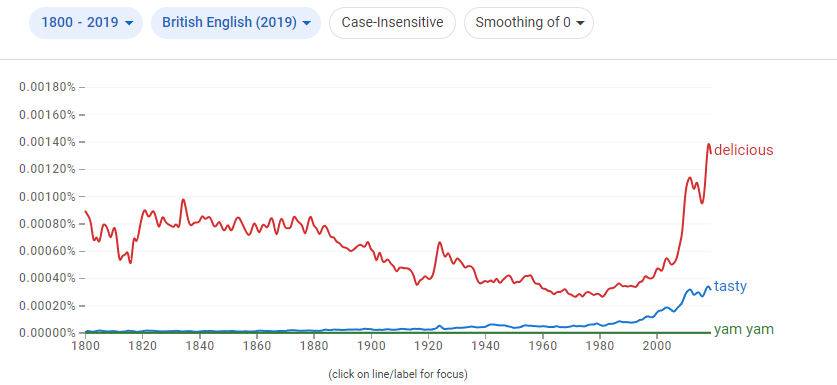






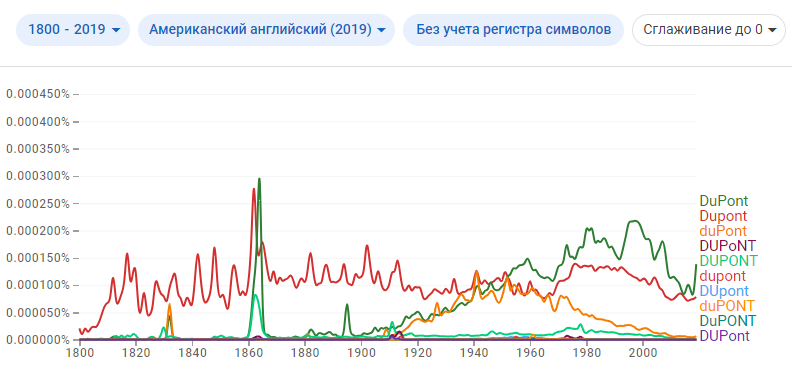
Smoothing





При сравнении частот использования слов "tasty", "delicious" и "yam yam" в корпусах британского и американского английского с 1800 по 2008 год с изменением параметра smoothing с 3 до 0 вид графика может измениться из-за различий в обработке данных.  
  
Параметр smoothing влияет на то, насколько сглаженными будут данные. С увеличением значения параметра smoothing график будет более сглаженным и менее подвержен шумам и выбросам. Наоборот, при уменьшении значения параметра smoothing данные будут менее сглаженными и более точно отражать фактические значения частот.  
  
Поэтому при изменении параметра smoothing с 3 до 0 можно ожидать более резких изменений на графике, что может привести к более выраженным пикам и провалам в зависимости от конкретных данных частот использования слов в корпусах.  
  
Сглаживание данных - это метод обработки данных, который помогает уменьшить шумы и выбросы в данных, делая их более однородными и легкими для анализа. Сглаживание используется для уменьшения колебаний в данных и выявления общих тенденций или трендов.  
  
Существует несколько методов сглаживания данных, включая скользящее среднее, экспоненциальное сглаживание, полиномиальное сглаживание и другие. Каждый метод имеет свои особенности и подходит для определенных типов данных и целей анализа.  
  
Сглаживание данных может быть полезным при работе с временными рядами, анализе трендов, прогнозировании и других задачах, где важно выделить общие закономерности и уменьшить влияние случайных колебаний.

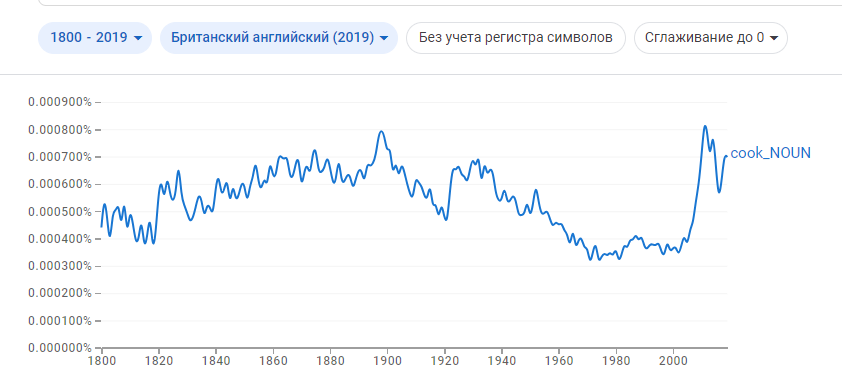
**2. **

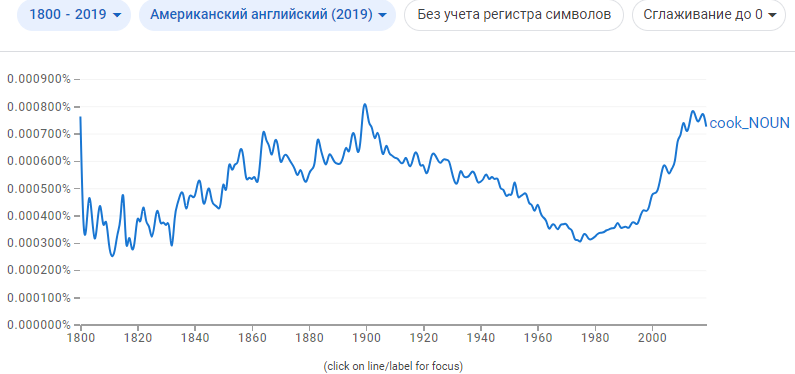
****

**3.**

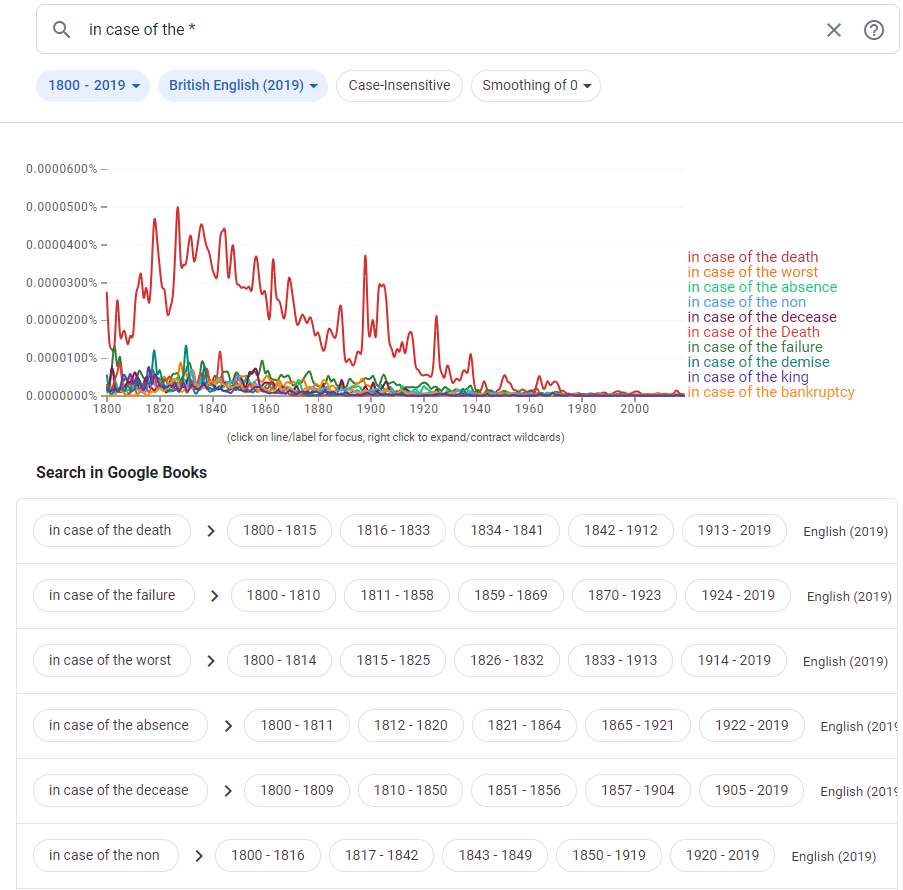
****

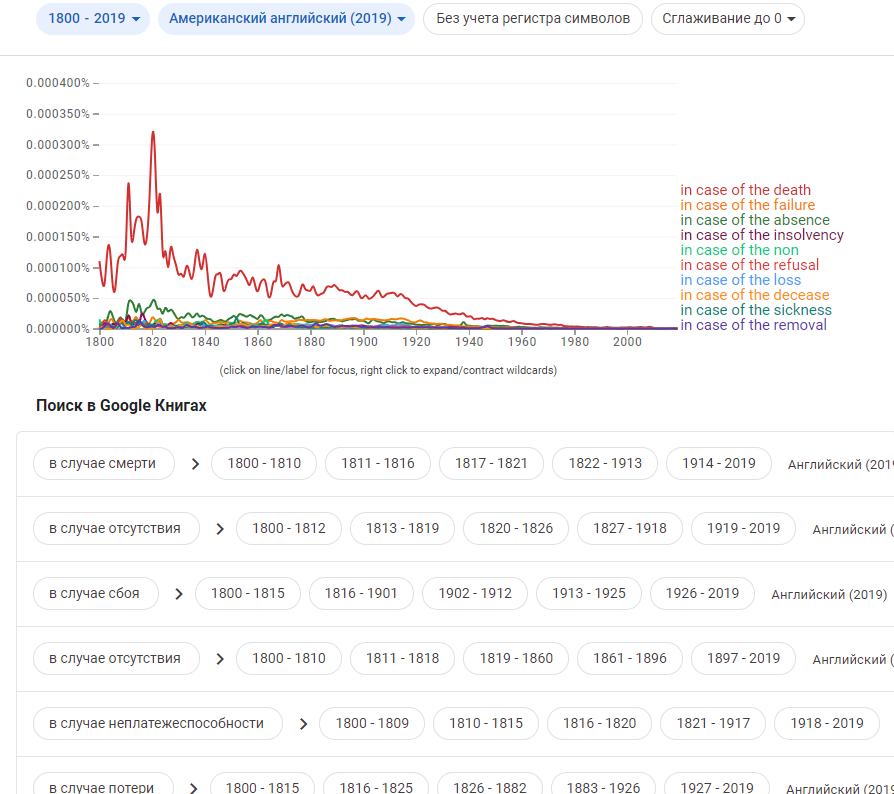
****

****

****

**4.**

****

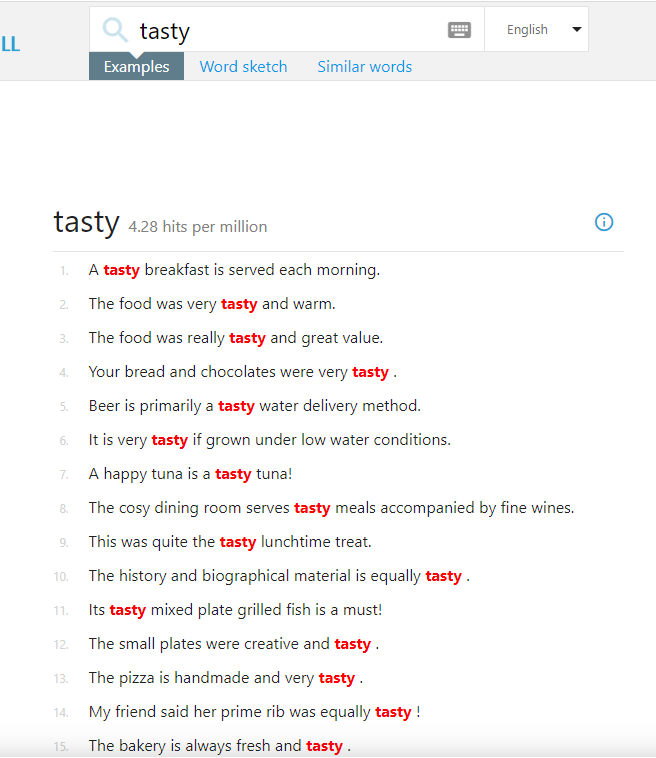
****

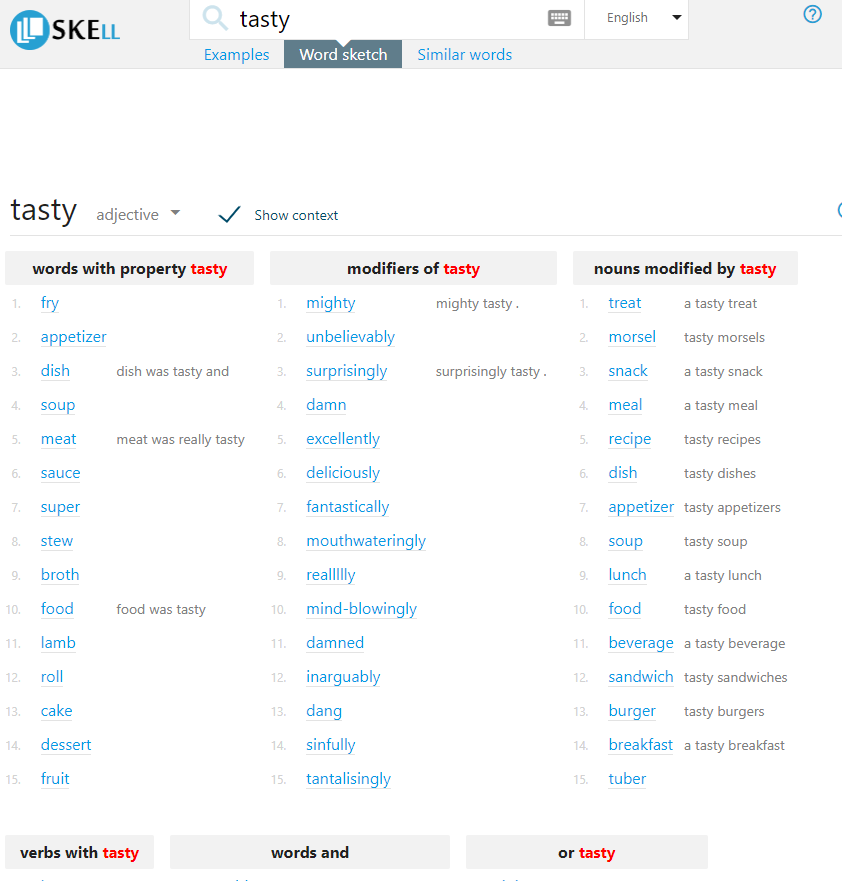
**5.**

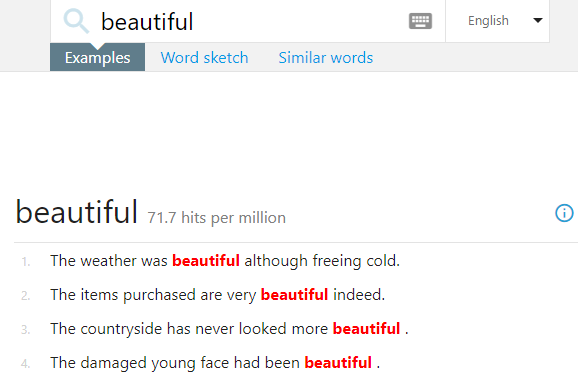
****

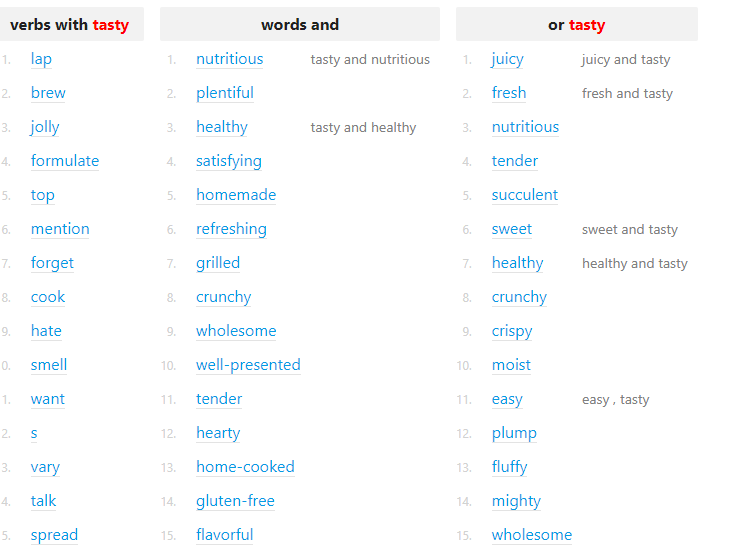
****

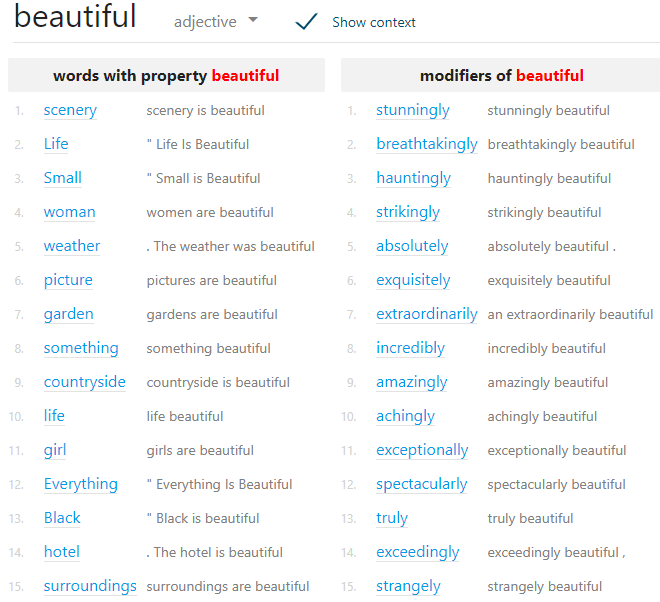
**Задания 2.1**

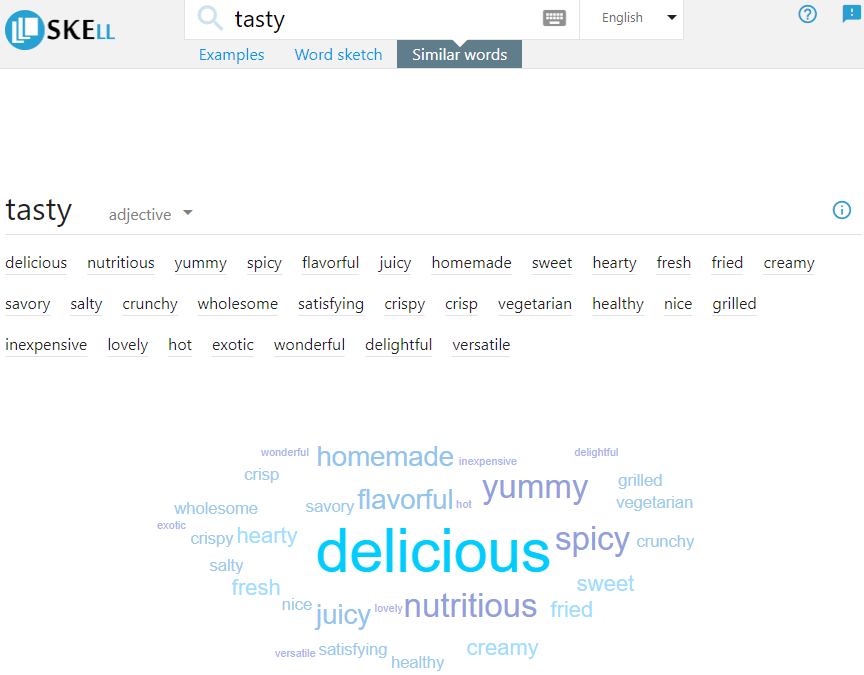
****

**2) **

****

****

****

****

****

1. Примеры:  
Инструмент "Примеры" на сайте SkeLL позволяет увидеть, как используется конкретное слово в контексте предложений. Это помогает понять правильный контекст использования слова и его значения. Возможные сферы применения: изучение новых слов, проверка правильности использования слова в контексте, анализ структуры предложений.  
  
2. Сочетаемость слов:  
Инструмент "Сочетаемость слов" позволяет увидеть, с какими другими словами часто сочетается данное слово. Это помогает понять, какие слова часто используются вместе с данным и какие фразы или выражения чаще всего встречаются. Возможные сферы применения: составление фраз и выражений, изучение коллокаций, анализ структуры текстов.  
  
3. Похожие слова:  
Инструмент "Похожие слова" позволяет найти синонимы, антонимы или другие слова, которые связаны с исходным словом по смыслу или тематике. Это помогает расширить словарный запас, найти подходящие замены или альтернативы при написании текстов. Возможные сферы применения: поиск синонимов и антонимов, расширение словарного запаса, улучшение стиля и качества текстов.  
  
Каждый из этих инструментов может быть полезен при изучении языка, написании текстов, анализе лингвистических данных и других областях, где требуется работа со словами и текстом.

**Что обозначают цифры**

**1.** Файлы, которые составляют наш корпус.

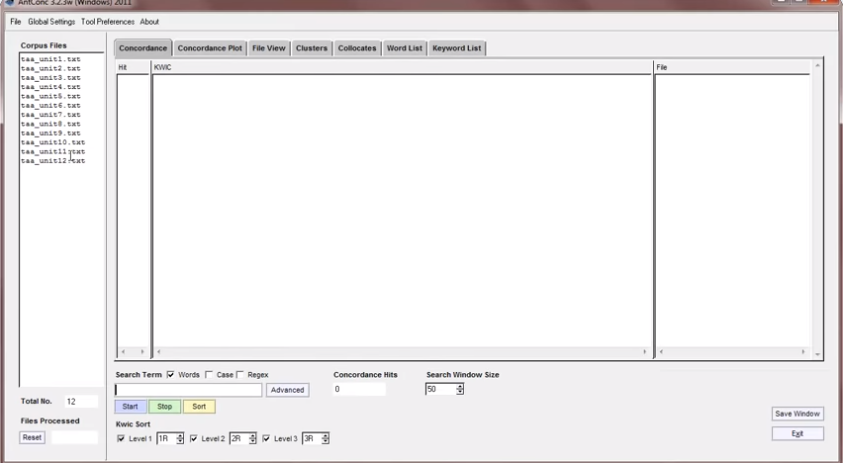
2.Word list - программа выведет список слов из книги, отсортированных по частоте.

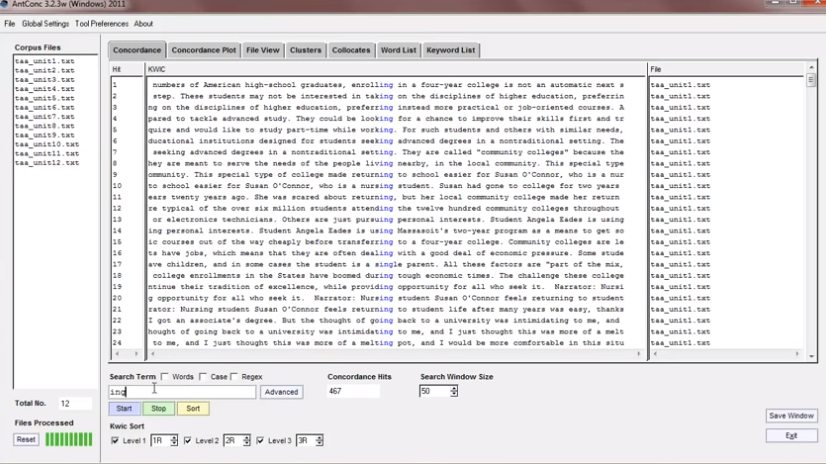
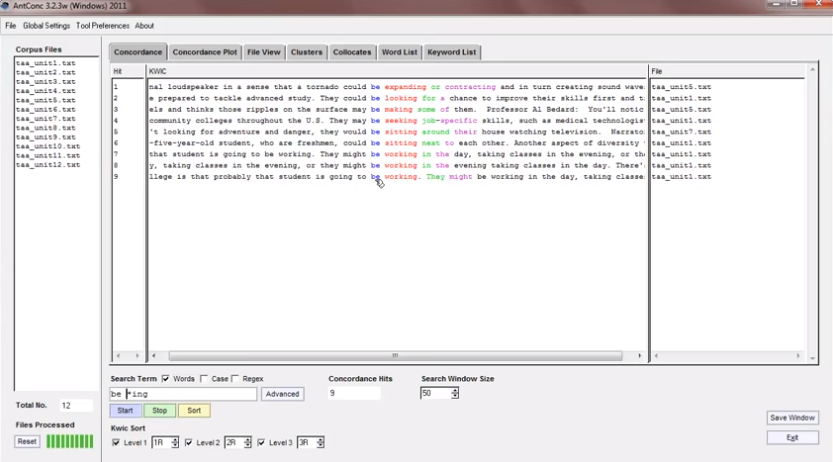
3. Конкорданс - это список всех употреблений заданного языкового выражения (например, слова) в контексте, возможно, со ссылками на источник.

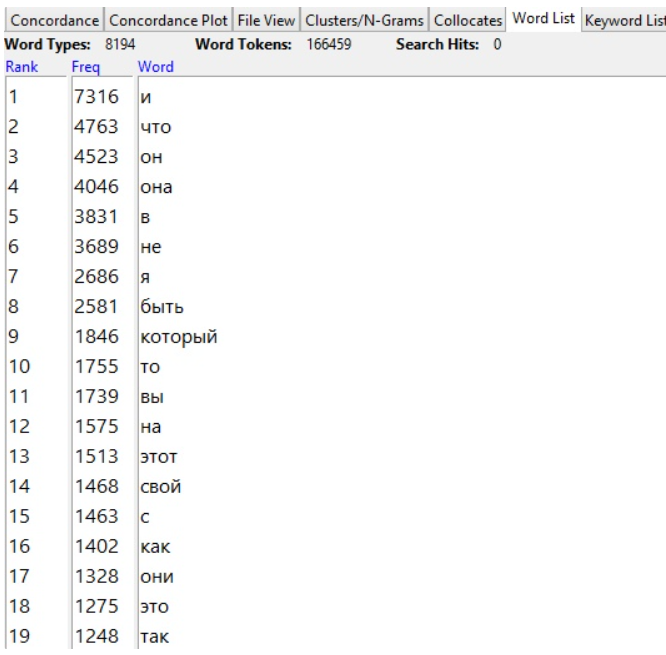
4. Кластеры (Clusters). Инструмент кластеры используется для создания упорядоченного списка кластеров, которые появляются вокруг поиска в целевом файле, перечисленные в левой части главного окна.

**Задания по AncConc**

1. Concordance (Конкорданс):  
   Инструмент "Concordance" в AntConc позволяет просмотреть все вхождения определенного слова или фразы в тексте, с контекстом и частотой использования. Это помогает анализировать употребление слова в различных контекстах, выявлять его частотность и использование. Возможные сферы применения: лингвистические исследования, анализ текстов на определенную тему, изучение структуры текстов.  
     
   2. Clusters (Кластеры):  
   Инструмент "Clusters" в AntConc позволяет выделить группы слов, которые часто встречаются вместе в тексте. Это помогает определить тематику текста, выделить ключевые термины и выявить связи между различными словами. Возможные сферы применения: анализ тематики текстов, изучение ключевых слов и понятий, выявление связей между словами.  
     
   3. Word list (Список слов):  
   Инструмент "Word list" в AntConc отображает список всех уникальных слов, используемых в тексте, с указанием их частоты встречаемости. Это помогает оценить словарный запас текста, выделить ключевые слова и провести анализ частотности слов. Возможные сферы применения: анализ словарного запаса текста, изучение употребления слов в контексте, подготовка к изучению нового языка.  
     
   4. Key words (Ключевые слова):  
   Инструмент "Key words" в AntConc выделяет наиболее значимые и часто употребляемые слова в тексте, помогая определить его основную тему или содержание. Это полезно для быстрой оценки содержания текста и выделения ключевых моментов. Возможные сферы применения: анализ содержания текста, выделение ключевых терминов, определение тематики.  
     
   Каждый из этих инструментов в AntConc может быть полезен при анализе текстов, проведении лингвистических исследований, изучении языка или работы с большим объемом текстовых данных.



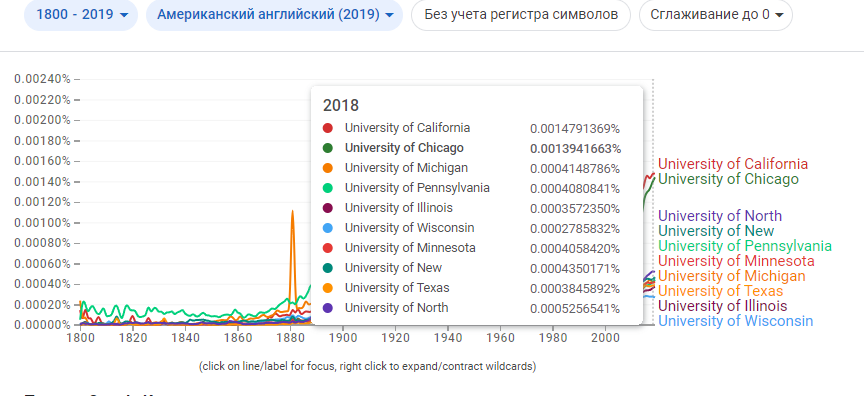
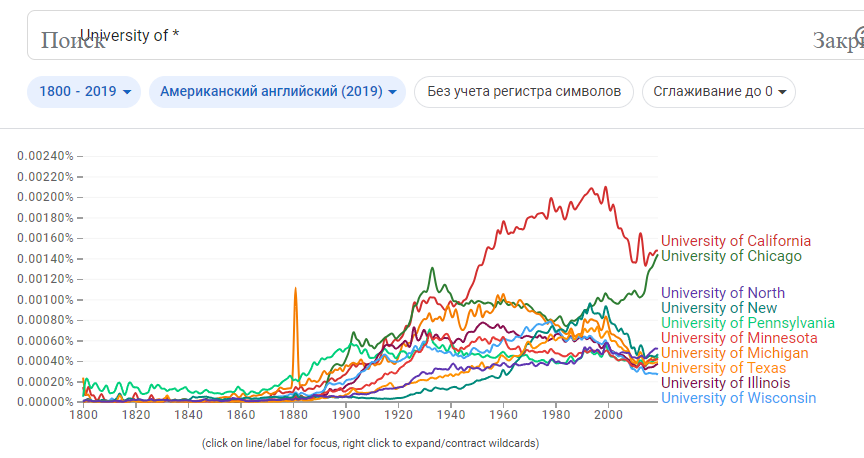
****

****

****

**Уровень 2**

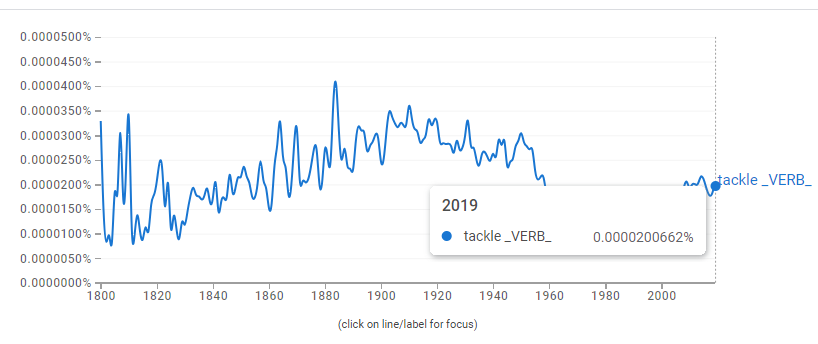
**Задание 6**

****

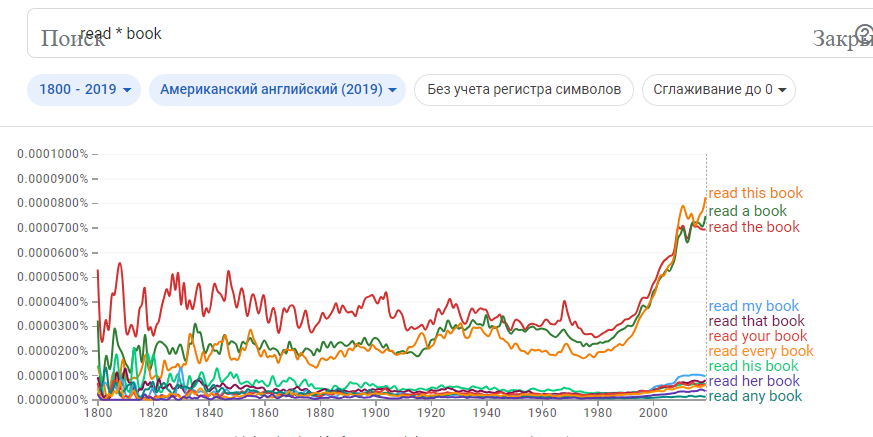
****

**Задание 7**

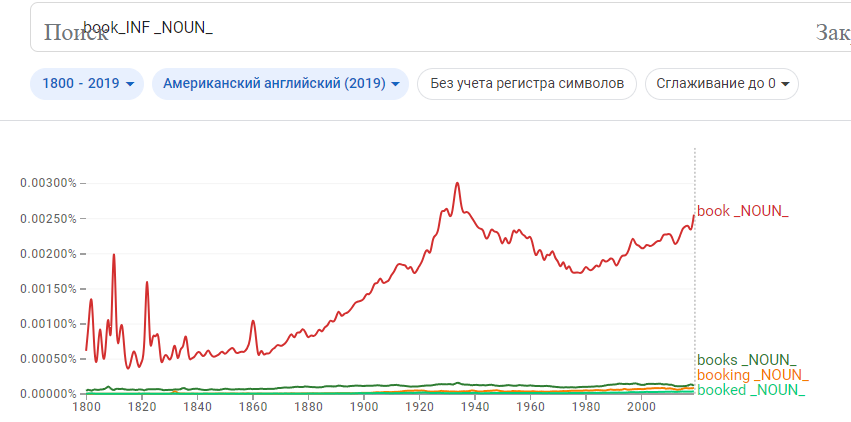
****

****

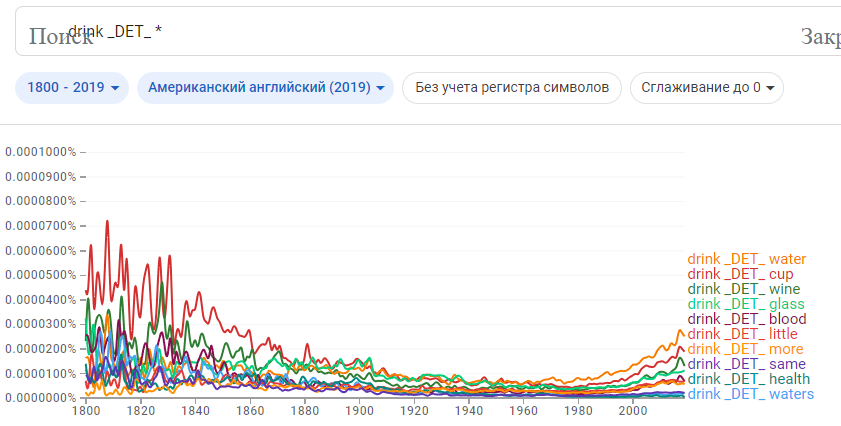
**Задание 8**

****

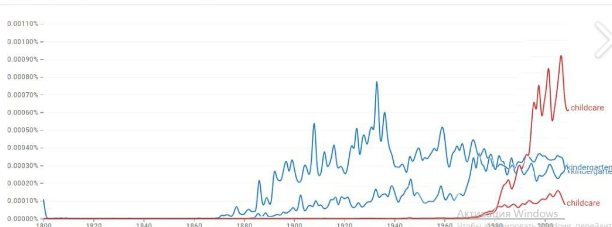
**Задание 9**

****

**Задание 10**

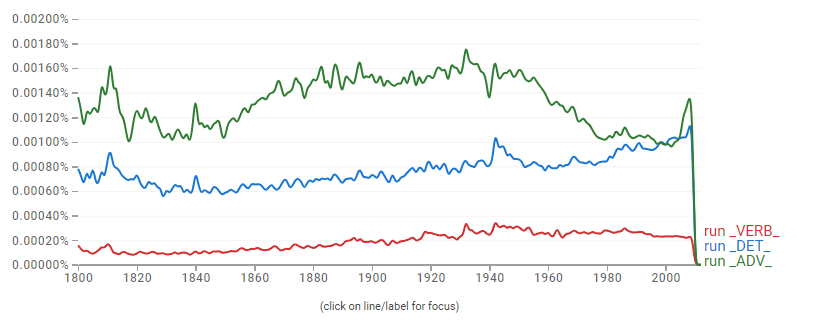
****

**Задание 11**

****

Американский

**Задание 12**

****

**Задание 13**

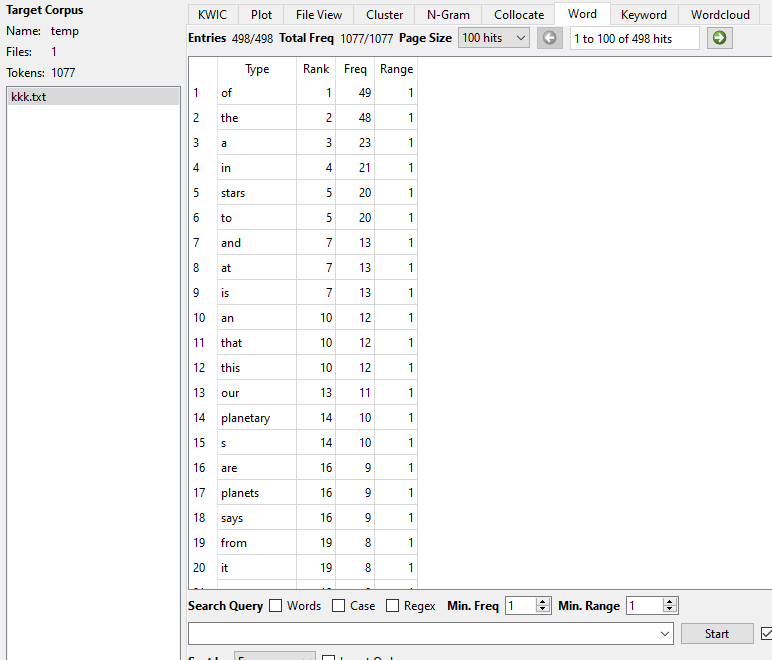
****

****

**Задание 2.2)**

1. Word Sketch (Ассоциации слов): Этот инструмент позволяет анализировать семантические и грамматические ассоциации для конкретного слова. Результаты отображаются в виде списка слов и фраз, которые часто встречаются в контексте данного слова. Это полезно для изучения значений и употребления слов в различных контекстах.  
     
   2. Word Sketch Difference (Различия в ассоциациях слов): Сравнивает ассоциации двух слов и выделяет различия в их употреблении. Это помогает понять семантические различия между словами и контексты, в которых они используются.  
     
   3. Trends (Тенденции): Показывает изменения в употреблении слова или фразы с течением времени. Это полезно для изучения эволюции языка и определения тенденций в употреблении конкретных выражений.  
     
   4. Thesaurus (Тезаурус): Предоставляет синонимы, антонимы и связанные слова для заданного слова. Это помогает расширить словарный запас и найти подходящие альтернативы в тексте.  
     
   5. Wordlist (Список слов): Показывает список слов, которые часто встречаются в контексте заданного слова. Это помогает понять употребление слова в различных контекстах.  
     
   6. Show Visualization (Визуализация): Позволяет визуализировать результаты анализа данных, например, через графики или диаграммы, что делает информацию более наглядной и понятной.  
     
   7. Concordance (Конкорданс): Дополнительные возможности этого раздела могут включать фильтрацию результатов по различным параметрам, сортировку по частоте употребления или алфавиту, а также возможность просмотра контекста вокруг заданного слова для лучшего понимания его использования.

Задание (Уровни 2, 3): Скачайте статью на английском. Загрузите в антконк, сделайте список частотных лемм.

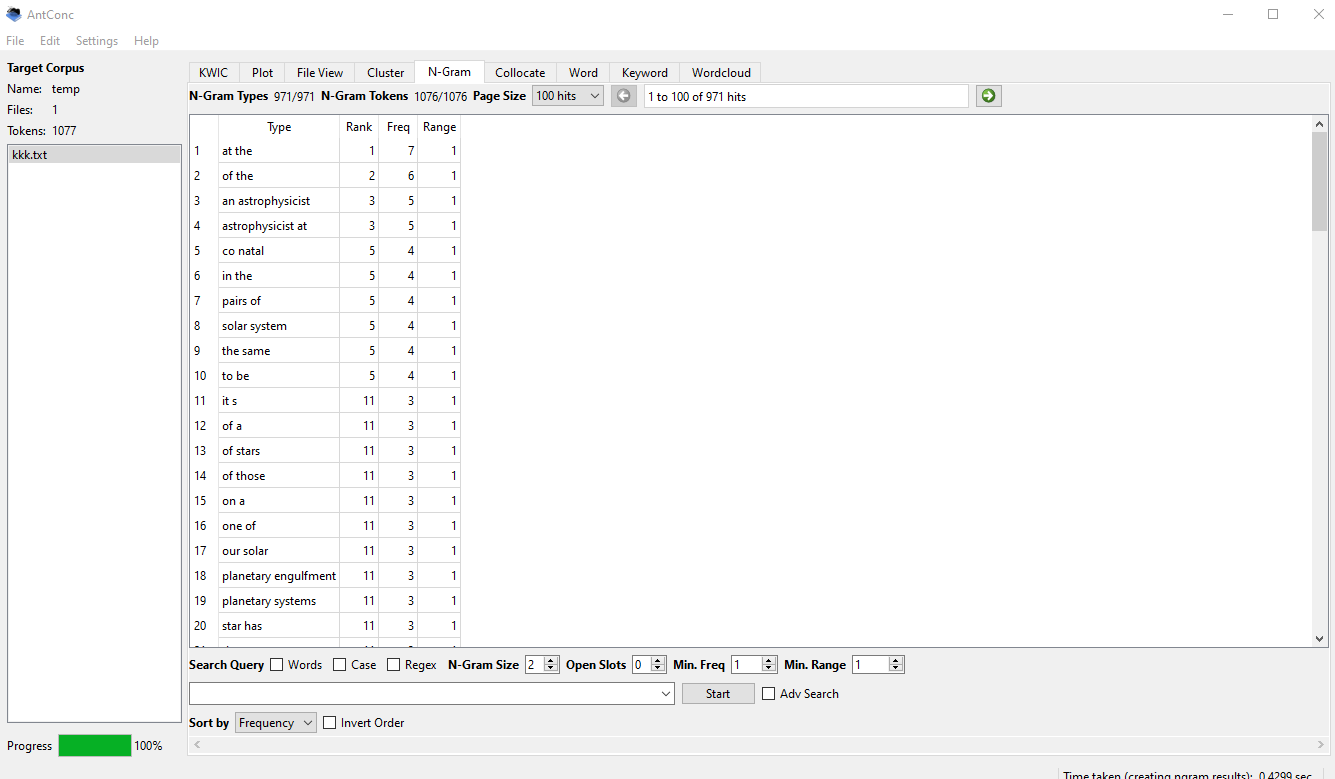
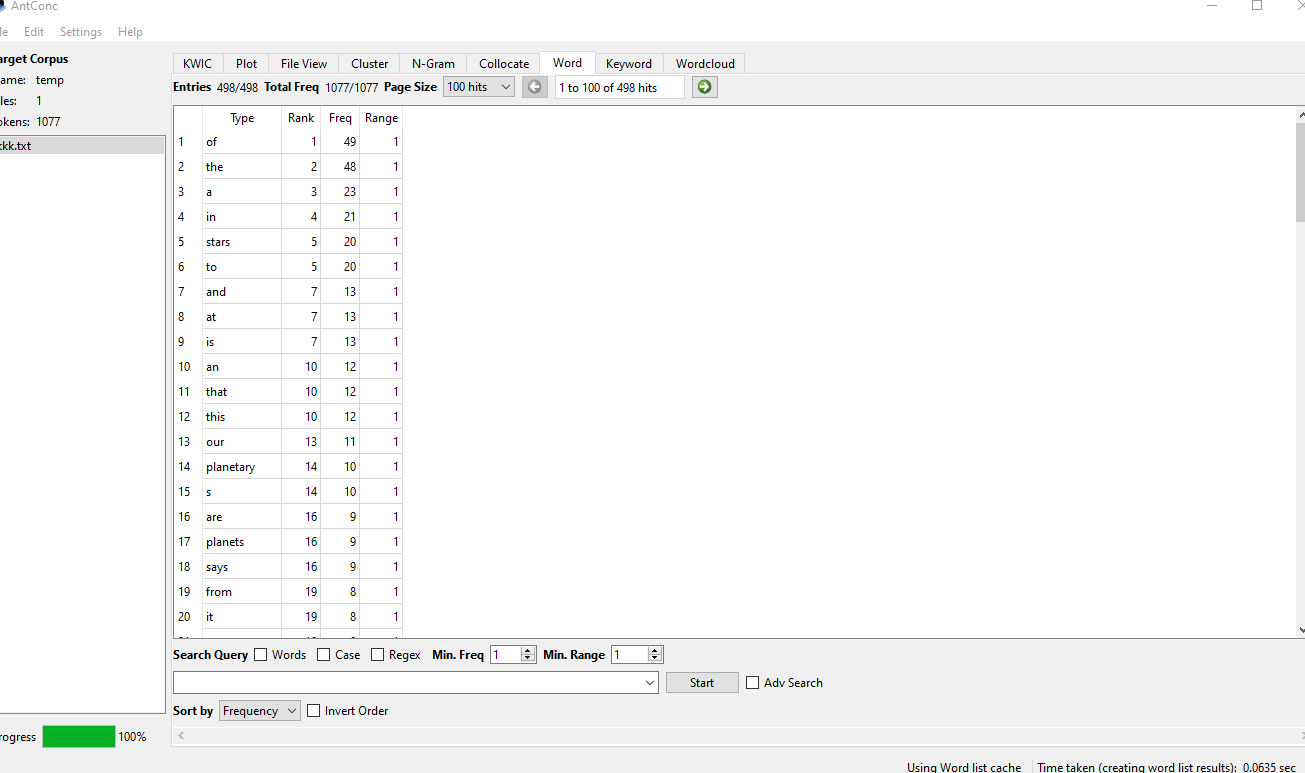
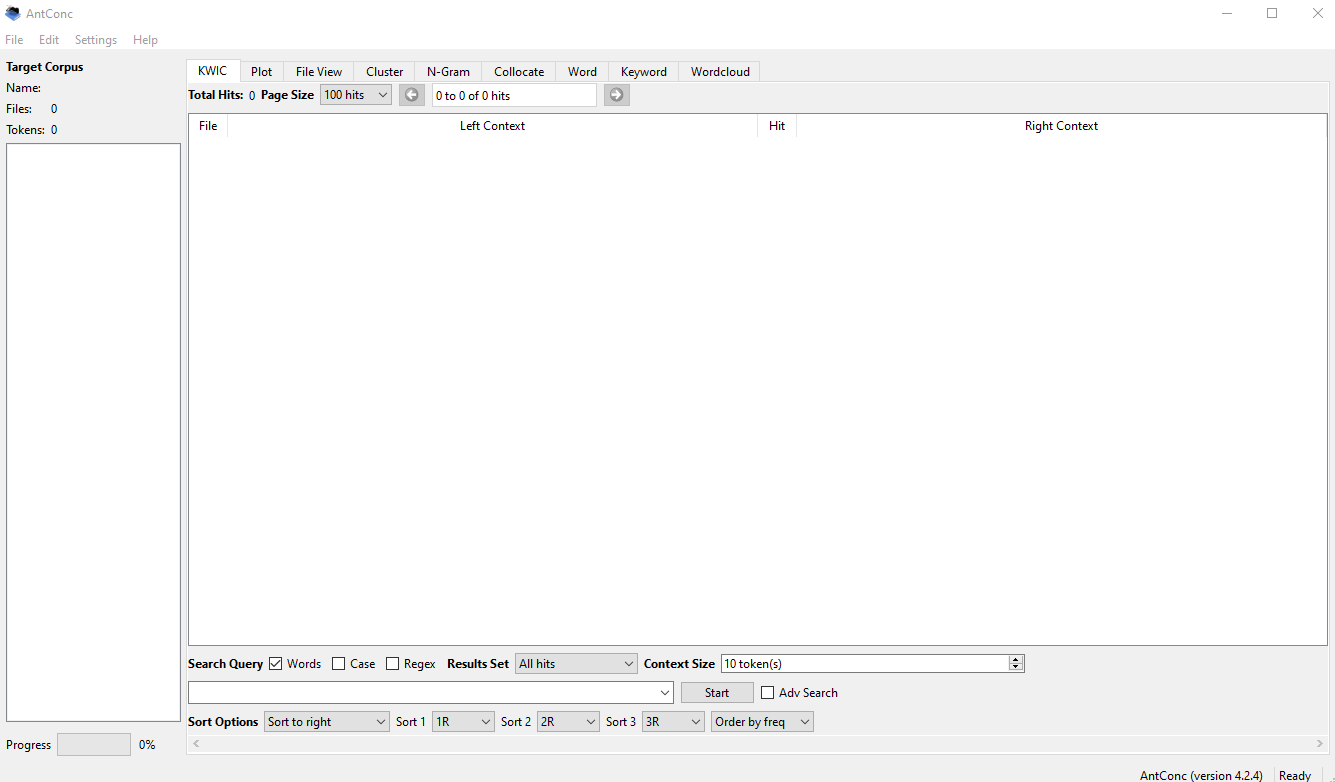
****

**Уровень 3**

**9 corpus software**

1. **AntConc**

AntConc - это программа для анализа текста, которая предоставляет различные инструменты и функции для работы с текстовыми данными. Некоторые из основных функций AntConc включают:  
  
1. Конкорданс - позволяет искать слова или фразы в тексте и отображать их контексты.  
  
2. Список слов - показывает частоту встречаемости слов в тексте.  
  
3. Коллокации - выявляет слова, которые часто встречаются вместе.  
  
4. Дисперсионный анализ - показывает, как распределены слова в тексте.  
  
5. Анализ ключевых слов - определяет наиболее значимые слова в тексте.  
  
6. Сегментация текста - разбивает текст на отдельные сегменты для анализа.  
  
AntConc обладает удобным пользовательским интерфейсом и может быть использован как для лингвистических исследований, так и для анализа текстов в образовательных целях.

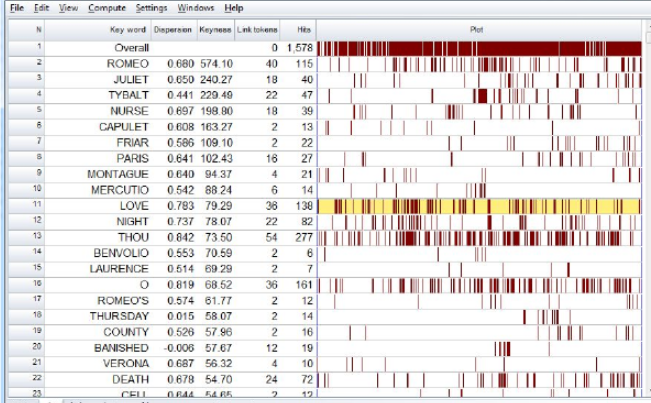
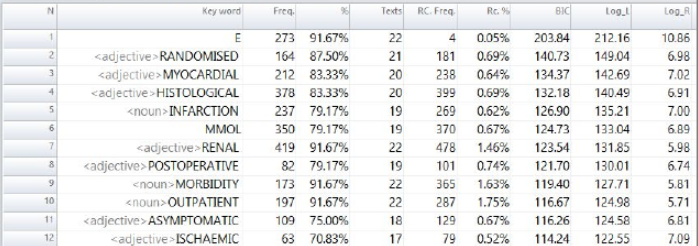


1. WMATRIX

Wmatrix - это еще один инструмент для анализа текста, который предоставляет различные функции для работы с текстовыми данными. Некоторые из основных функций Wmatrix включают:  
  
1. Анализ лексического разнообразия - позволяет оценить уровень разнообразия слов и лексическую насыщенность текста.  
  
2. Анализ ключевых слов - выделяет наиболее значимые слова в тексте.  
  
3. Коллокации - выявляет слова, которые часто встречаются вместе и могут иметь особое значение.  
  
4. Анализ структуры предложений - позволяет исследовать структуру предложений в тексте.  
  
5. Построение графиков и диаграмм - для визуализации результатов анализа текста.  
  
6. Интерфейс для работы с большими объемами текстовых данных.  
  
Wmatrix также предоставляет удобный пользовательский интерфейс и может быть использован для лингвистических исследований, анализа корпусов текстов, а также для обучения и преподавания языков.

1. WORDSMITH TOOLS

Wordsmith Tools - это комплексное программное обеспечение для анализа текста и лингвистических исследований. Вот основные функции и возможности Wordsmith Tools:  
  
1. Частотный анализ: Программа позволяет проводить частотный анализ текста, выявляя наиболее часто встречающиеся слова, биграммы, триграммы и другие языковые единицы. Это помогает выявить ключевые слова и фразы, а также понять частоту их использования.  
  
2. Конкордансный анализ: Wordsmith Tools предоставляет возможность создания конкордансов - списка контекстов, в которых встречается определенное слово или фраза. Это помогает изучать употребление слова в различных контекстах.  
  
3. Анализ коллокаций: Программа позволяет проводить анализ коллокаций - часто встречающихся сочетаний слов или фраз. Это помогает выявить связи между словами и понять их семантическую структуру.  
  
4. Статистический анализ: Wordsmith Tools предоставляет возможность проводить различные статистические анализы текста, такие как распределение слов по частоте, длине и другим параметрам.  
  
5. Инструменты для работы с корпусами текстов: Программа позволяет создавать и работать с корпусами текстовых данных, проводить сравнительный анализ текстов и извлекать информацию из больших объемов текста.  
  
6. Визуализация данных: Wordsmith Tools предоставляет возможность визуализации результатов анализа текста с помощью графиков, диаграмм и других визуальных элементов.  
  
7. Интерфейс для работы с различными языками: Программа поддерживает работу с различными языками и позволяет проводить анализ текста на разных языках.  
  
Wordsmith Tools является мощным инструментом для лингвистических исследований, анализа текстовых данных, компьютерной лингвистики и других областей, где требуется работа с текстом и его анализ.

** **

1. SKETCH ENGINE

Sketch Engine - это инструмент для работы с корпусами текстов, который предоставляет широкий спектр функций и возможностей для лингвистических исследований. Вот основные функции и возможности Sketch Engine:  
  
1. Создание и управление корпусами текстов: Sketch Engine позволяет создавать и управлять корпусами текстовых данных различных языков и типов. Пользователи могут загружать свои собственные тексты или использовать готовые корпуса из библиотеки Sketch Engine.  
  
2. Лемматизация и морфологический анализ: Программа предоставляет возможность проводить лемматизацию текстов и морфологический анализ слов, что позволяет выявлять формы слов, их грамматические характеристики и связи.  
  
3. Частотный анализ и статистика: Sketch Engine позволяет проводить частотный анализ текста, выявляя наиболее часто встречающиеся слова, коллокации, биграммы, триграммы и другие языковые единицы. Также доступны различные статистические инструменты для анализа текстов.  
  
4. Конкордансный анализ: Программа предоставляет возможность создания конкордансов - списка контекстов, в которых встречается определенное слово или фраза. Это помогает изучать употребление слова в различных контекстах.  
  
5. Анализ коллокаций и семантических связей: Sketch Engine позволяет проводить анализ коллокаций - часто встречающихся сочетаний слов или фраз, а также изучать семантические связи между словами.  
  
6. Визуализация данных: Программа предоставляет возможность визуализации результатов анализа текста с помощью графиков, диаграмм и других визуальных элементов.  
  
7. Инструменты для работы с параллельными корпусами: Sketch Engine поддерживает работу с параллельными корпусами текстов на разных языках, что позволяет проводить сравнительный анализ и исследования перевода.  
  
Sketch Engine является мощным инструментом для лингвистических исследований, компьютерной лингвистики, изучения языков и других областей, где требуется работа с текстом и его анализ.



1. CORPUS.BYU.EDU

[Corpus.byu.edu](https://vk.com/away.php?utf=1&to=http%3A%2F%2FCorpus.byu.edu) - это онлайн-платформа, предоставляющая доступ к большому количеству корпусов текстов на различных языках. Вот основные функциональные возможности [Corpus.byu.edu](https://vk.com/away.php?utf=1&to=http%3A%2F%2FCorpus.byu.edu):  
  
1. Доступ к разнообразным корпусам текстов: Пользователям предоставляется доступ к различным корпусам текстов на разных языках, включая английский, испанский, китайский и др. Корпуса содержат тексты различных жанров и стилей, что позволяет проводить широкий спектр лингвистических исследований.  
  
2. Поиск и фильтрация данных: Пользователи могут осуществлять поиск по корпусам текстов с использованием различных параметров, таких как ключевые слова, части речи, год публикации и другие. Также имеется возможность фильтровать данные по различным критериям.  
  
3. Частотный анализ: Платформа предоставляет инструменты для проведения частотного анализа текстов. Пользователи могут определить наиболее часто встречающиеся слова, выражения, коллокации и другие языковые единицы в выбранном корпусе.  
  
4. Конкордансный анализ: [Corpus.byu.edu](https://vk.com/away.php?utf=1&to=http%3A%2F%2FCorpus.byu.edu) позволяет создавать конкордансы - списки контекстов, в которых встречается определенное слово или фраза. Это помогает изучать употребление слова в различных контекстах.  
  
5. Сравнительный анализ: Пользователи могут сравнивать данные из разных корпусов текстов или проводить анализ текстов на разных языках для изучения сходств и различий.  
  
6. Инструменты для исследования языковых явлений: [Corpus.byu.edu](https://vk.com/away.php?utf=1&to=http%3A%2F%2FCorpus.byu.edu) предоставляет инструменты для исследования различных языковых явлений, таких как синтаксис, семантика, стилистика и другие.  
  
7. Экспорт данных: Пользователи могут экспортировать результаты анализа текстов в различные форматы для дальнейшего использования или обработки.  
  
[Corpus.byu.edu](https://vk.com/away.php?utf=1&to=http%3A%2F%2FCorpus.byu.edu) представляет собой полезный инструмент для лингвистических исследований, обучения языкам, компьютерной лингвистики и других областей, требующих работы с корпусами текстов.

****

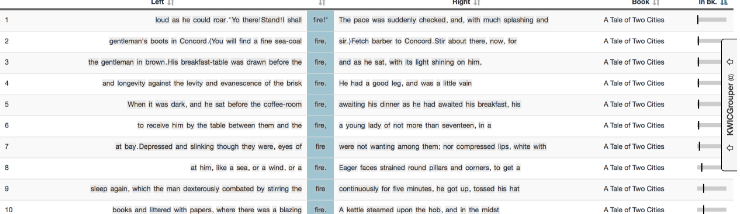
1. CQP WEB

CQPweb (Corpus Query Processor web) - это онлайн-интерфейс для работы с корпусами текстов, который предоставляет возможности по выполнению сложных поисковых запросов и анализу текстовых данных. Вот основные функциональные возможности CQPweb:  
  
1. Поиск по корпусам: Пользователи могут осуществлять поиск по различным корпусам текстов, используя сложные лингвистические запросы. Они могут задавать условия, такие как слова, части речи, синтаксические конструкции и другие параметры для поиска нужной информации.  
  
2. Анализ результатов поиска: CQPweb предоставляет инструменты для анализа результатов поиска, включая возможность просмотра контекстов, конкордансов, частотных списков и других статистических данных.  
  
3. Создание сложных запросов: Пользователи могут создавать сложные запросы, используя лингвистические операторы, регулярные выражения и другие инструменты для точного и гибкого поиска в текстах.  
  
4. Сравнительный анализ: CQPweb позволяет сравнивать данные из разных корпусов текстов или проводить сравнительный анализ текстов на разных языках для изучения различий и сходств.  
  
5. Интеграция с другими инструментами: CQPweb может интегрироваться с другими инструментами для лингвистического анализа, что позволяет пользователям проводить более широкий спектр исследований.  
  
6. Экспорт данных: Пользователи могут экспортировать результаты анализа в различные форматы для дальнейшего использования или обработки.  
  
CQPweb представляет собой мощный инструмент для лингвистических исследований, обучения языкам, компьютерной лингвистики и других областей, требующих работы с корпусами текстов.

****

1. CLIC (CORPUS LINGUISTICS IN CONTEXT)

Clic - это программное обеспечение для лингвистического анализа текстов, которое предоставляет различные функциональные возможности для работы с корпусами. Вот основные функции Clic:  
  
1. Импорт и экспорт данных: Clic позволяет импортировать текстовые данные из различных источников, таких как файлы текстов, базы данных и др. Также есть возможность экспорта результатов анализа в различные форматы.  
  
2. Поиск и фильтрация: Пользователи могут выполнять поиск по текстовым данным с помощью различных параметров, таких как слова, фразы, части речи и другие лингвистические характеристики. Также можно применять фильтры для уточнения результатов.  
  
3. Конкорданс: Clic предоставляет возможность создания конкордансов - списков контекстов, в которых встречается искомое слово или выражение, с возможностью просмотра контекстов.  
  
4. Статистический анализ: С помощью Clic можно проводить статистический анализ текстов, включая подсчет частотности слов, коллокаций, распределения частей речи и других лингвистических параметров.  
  
5. Визуализация данных: Пользователи могут визуализировать результаты анализа с помощью графиков, диаграмм и других инструментов для наглядного представления информации.  
  
6. Работа с различными языками: Clic поддерживает работу с текстами на разных языках, что делает его универсальным инструментом для лингвистических исследований.  
  
7. Интерфейс пользователя: Интерфейс Clic дружелюбен к пользователю и обеспечивает удобство в работе с текстовыми данными и результатами анализа.  
  
Clic является полезным инструментом для лингвистических исследований, обучения языкам, компьютерной лингвистики и других областей, где требуется работа с корпусами текстов.

** **

1. TextSTAT

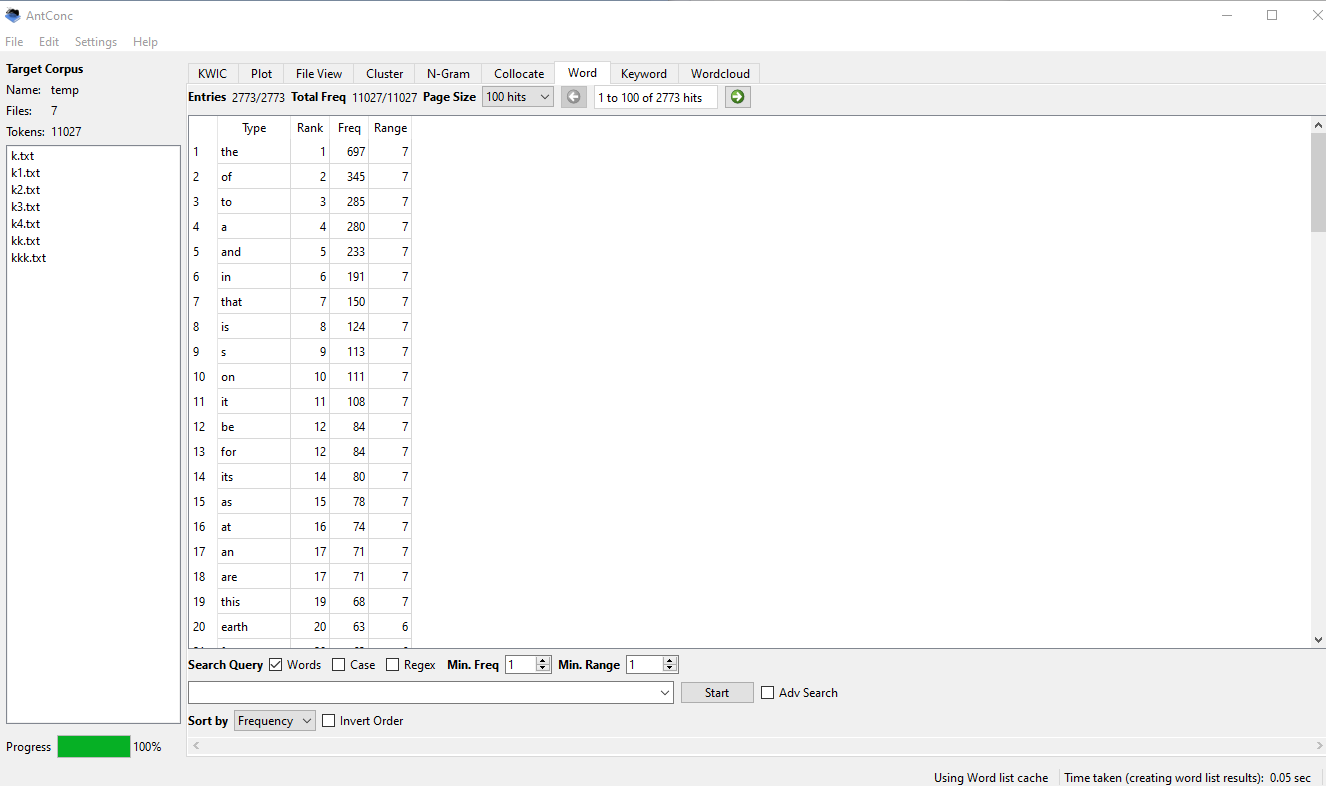
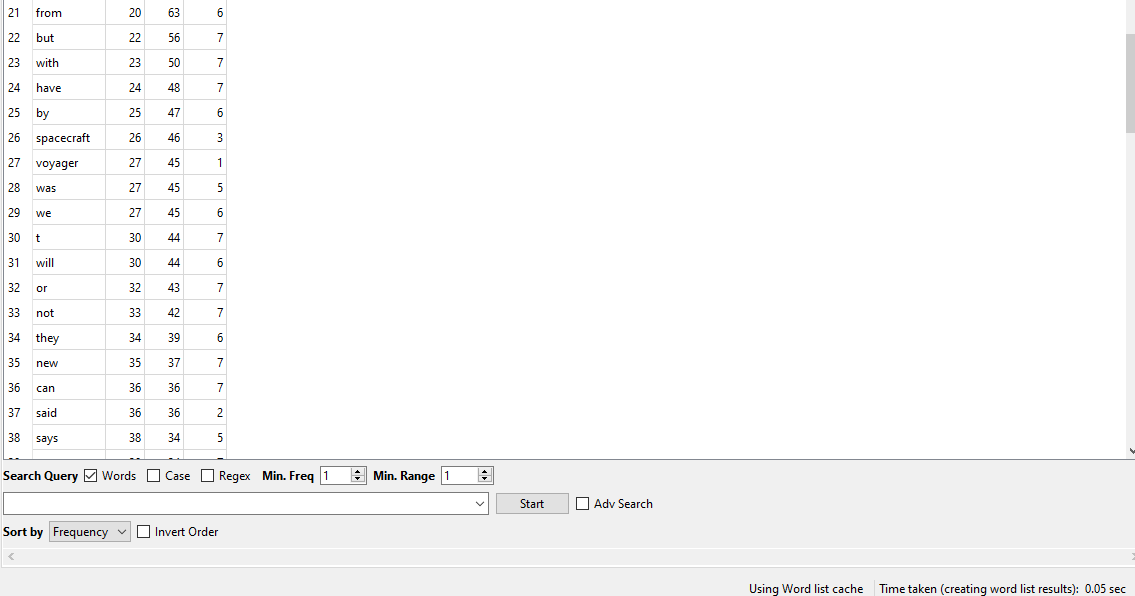
Textstart - это инструмент для лингвистического анализа текстов, который предоставляет разнообразные функциональные возможности для работы с корпусами текстов. Вот основные функции Textstart:  
  
1. Импорт и экспорт данных: Textstart позволяет импортировать текстовые данные из различных источников, таких как файлы текстов, базы данных, веб-страницы и другие. Также есть возможность экспорта результатов анализа в различные форматы.  
  
2. Токенизация и лемматизация: Textstart предоставляет возможность токенизации текстов (разделение на отдельные слова или токены) и лемматизации (приведение слов к их базовой форме).  
  
3. Частеречная разметка: С помощью Textstart можно проводить частеречную разметку текстов, то есть определение частей речи для каждого слова.  
  
4. Анализ частотности: Инструмент позволяет проводить анализ частотности слов, выявлять самые часто встречающиеся слова в тексте или корпусе.  
  
5. Коллокации и ассоциации: Textstart помогает выявлять коллокации - слова, которые часто встречаются вместе, а также ассоциации между словами.  
  
6. Визуализация данных: Пользователи могут визуализировать результаты анализа с помощью графиков, облаков слов, диаграмм и других инструментов для наглядного представления информации.  
  
7. Интерфейс пользователя: Textstart обладает удобным интерфейсом, который облегчает работу с текстовыми данными и результатами анализа.  
  
Textstart является полезным инструментом для лингвистических исследований, обучения языкам, анализа текстов на различных языках и других задач, связанных с обработкой текстовых данных.

**Лабораторная работа № 9 «Корпусная лингвистика. Корпусные менеджеры».docx**

Для выполнения данного задания было принято решение использовать научные статьи (т.к. таким способом легче собрать нужный объем корпуса) про космос (освоение космоса, интересные факты про космос и т.д.), т.к. это одна из самых волнующих человечество тем последнего столетия. Количество словоупотреблений в полученном корпусе составляет 11.027 слов, что удовлетворяет условию задания. Тексты были отобраны на просторах сети Интернет. Критерий отбора: тексты, в которых термины могут иметь несколько значений. Авторами статей были мужчины, носители английского языка за 50 лет, ученые, преимущественно американцы, даты выхода статей – 2009-2022 годы.

Цель исследования: выявить, какие слова чаще всего встречаются в научных текстах про космос, с какими словами обычно употребляется слово space, solar system, planet и т.д., вычленить значение научных терминов из контекста.

1. Просмотр самых частотных слов в корпусе

Самыми частотными словами были (если исключить артикли, предлоги, местоимения и т.д., что мешает анализу) earth(63), spacecraft(46), voyager(45), new(37), space(35), says(34), can(36), moon(29), sun(29), system(29), time(27), solar(25), launch(24), planet(21)

Слово Earth чаще всего употребляется в сочетании со словами:



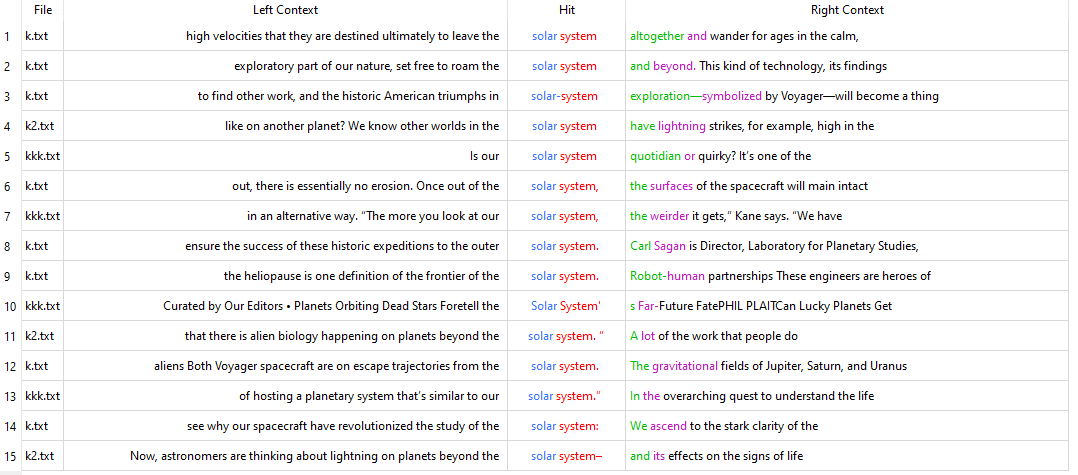
Слово sun чаще всего употребляется в сочетании со словами:



Слово planet чаще всего употребляется в сочетании со словами:



Слово solar system чаще всего употребляется в сочетании со словами:



Слово moon чаще всего употребляется в сочетании со словами:



Слово voyager чаще всего употребляется в сочетании со словами:

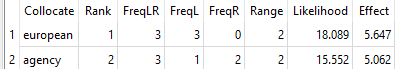


Слово space чаще всего употребляется в сочетании со словами:

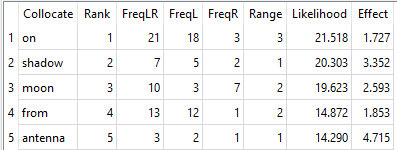


Посмотрим на коллокации:

Space



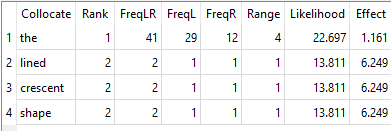
Earth



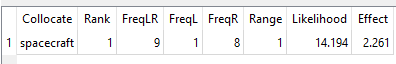
Planet



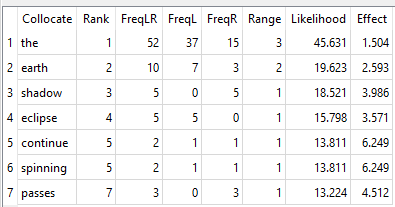
Sun



Voyager



Moon

****

**Вывод: таким образом, мы видим, что чаще всего в научных статьях про космос употребляются слова earth(63), spacecraft(46), voyager(45), new(37), space(35), says(34), can(36), moon(29), sun(29), system(29), time(27), solar(25), launch(24), planet(21). С помощью программы мы выявили, с какими словами чаще употребляются выше перечисленные, на основе этого можно сделать вывод: planet – речь о небесном теле, вращающимся по орбите вокруг звезды или её остатков, Earth – речь о нашей планете, Sun – речь о солнце, вокруг которого вращается наша планета и т.д.Проведенное исследование может помочь в более глобальных лингвистических исследованиях, в обучении языкам и других областях, в которых требуется работа с корпусами.**