Задание 1.

«Лексикограф. Глагол»

Ресурс "Лексикограф. Глагол" является онлайн-словарем, специализирующимся на глаголах русского языка. Он предоставляет пользователям возможность быстро и удобно найти информацию о значении, употреблении и особенностях употребления глагола. На сайте представлены различные разделы, которые помогают пользователю более полно и глубоко понять выбранный глагол.

В основе работы ресурса лежит обширный словарный запас, который позволяет найти нужную информацию о любом глаголе. Пользователь может найти определение глагола, узнать его синонимы и антонимы, увидеть примеры употребления в различных контекстах, изучить спряжение глагола в разных временах и лицах, а также узнать о специфических правилах использования конкретного глагола.

"Лексикограф. Глагол" обеспечивает доступ к качественной и проверенной лингвистической информации о глаголах, что делает его полезным инструментом для изучения и использования русского языка. Структурированная форма представления информации и удобный интерфейс делают работу с ресурсом эффективной и удобной для пользователей всех уровней знания русского языка.

“eLIBRARY.ru”

“eLIBRARY.ru” - это ресурс, предоставляющий доступ к электронным книгам, научным статьям, журналам и другим образовательным материалам. Здесь можно найти как платные, так и бесплатные публикации по различным областям знания: от науки и техники до искусства и гуманитарных наук.

На сайте “eLIBRARY.ru” вы можете оформить подписку на доступ ко всем материалам, либо приобрести отдельные книги или статьи. Ресурс также предоставляет возможность поиска информации по ключевым словам, авторам или темам, что упрощает поиск нужной литературы.

Кроме того, “eLIBRARY.ru” предоставляет возможность загрузки материалов для удобного чтения в офлайн-режиме, а также возможность сохранения интересующих вас публикаций в личном кабинете.

Таким образом, “eLIBRARY.ru” является ценным ресурсом для всех, кто интересуется образованием, наукой и культурой, обеспечивая доступ к широкому спектру образовательных и научных материалов.

“Этнология”

База данных "Этнология" является электронным ресурсом, который собирает, хранит и предоставляет информацию о различных этнических группах и культурах по всему миру. В этой базе данных можно найти разнообразные данные, связанные с этническими и культурными характеристиками различных народов, их обычаями, языками, религией, проживанием, традициями и многим другим.

Работа базы данных "Этнология" включает в себя сбор, хранение и актуализацию информации о различных культурах и этнических группах. Пользователи могут использовать этот ресурс для изучения и анализа информации о различных культурах, проведения исследований, подготовки учебных материалов, написания статей и многих других целей.

В целом, база данных "Этнология" играет важную роль в сохранении и продвижении знаний о разнообразии и культурном наследии различных народов по всему миру.

“Грамота.ру”

Портал Грамота.ру - это образовательный онлайн-ресурс, предназначенный для помощи людям в изучении русского языка, правилам русской орфографии и пунктуации, а также решения различных языковых вопросов.

На портале Грамота.ру можно найти следующие разделы и возможности:

1. Онлайн-справочник по правилам русского языка, где представлена информация о правилах орфографии, пунктуации, морфологии, синтаксиса и других языковых аспектах.

2. Грамотей - функция проверки правописания слов и фраз, автоматическое исправление ошибок и предложение верных вариантов написания.

3. Форум, где пользователи могут обсуждать языковые вопросы, задавать вопросы и давать советы другим.

4. Разделы "Школьникам" и "Студентам" с полезными материалами и учебными пособиями по русскому языку.

5. Игры и тесты для тренировки правописания и языковых навыков.

Работа портала Грамота.ру направлена на помощь пользователям в обучении и совершенствовании навыков русского языка, а также на повышение уровня грамотности и языковой культуры. Сайт является удобным и доступным ресурсом для всех, кто интересуется языком и желает улучшить свои языковые навыки.

“ИНИОН”

Сайт ИНИОН (Интеграционное информационное обеспечение национального хозяйства) является информационно-аналитическим порталом, предоставляющим широкий спектр данных и аналитики по различным аспектам экономической и социальной жизни России.

На сайте ИНИОН можно найти информацию о социо-экономической ситуации в стране, данные об инвестициях, финансах, торговле, промышленности, трудовом рынке, науке и образовании. Также здесь публикуются аналитические обзоры, отчеты, статистические данные, прогнозы и прочие материалы, позволяющие оценить текущее состояние экономики и принимать обоснованные решения.

Сайт ИНИОН предназначен для широкого круга пользователей – от государственных структур и бизнеса до обычных граждан, интересующихся экономической и социальной политикой России. Пользователи могут получить актуальную информацию, провести анализ рыночной ситуации и прогнозировать развитие отраслей экономики на основе представленных данных на сайте.

Задание 2.

1) Единицей корпуса в лингвистике может быть отдельное слово, предложение, текст, речь или любой другой языковой элемент, используемый для анализа и исследования языка. Количество и характер единиц корпуса зависят от конкретных целей исследования.

2) Отбор текстов для корпуса зависит от целей исследования, специфики корпуса и доступных ресурсов. Принципы отбора текстов могут быть разнообразными, но некоторые общие принципы включают:

1. Представительность: тексты должны отражать разнообразие жанров, тематик, стилей и диалектов, характерных для конкретного языка или языковой группы.

2. Равномерное покрытие: тексты должны покрывать различные периоды времени, географические области, социокультурные контексты и другие важные аспекты для обеспечения достоверности результатов анализа.

3. Сбалансированность: тексты должны быть равномерно распределены по различным категориям (например, по жанрам, авторам, тематикам) для представления всего многообразия языка.

Примером может служить Брауновский корпус (Brown Corpus), составленный в середине 20 века для английского языка. Тексты включают различные жанры, такие как проза, научные тексты, публицистика, разговоры и другие, чтобы обеспечить представительность стилей и тематик.

3) 1. Исследовательский корпус - это набор текстов, собранных и скомпилированных исследователем для специфической цели исследования. Этот тип корпуса может включать в себя тексты на определенную тему, жанр или язык, собранные и оцифрованные исследователем для анализа. Исследовательский корпус может быть создан с помощью различных методов, таких как веб-скрапинг или запросы к базам данных.

2. Статический корпус - это набор текстов, который не обновляется и не изменяется. Этот тип корпуса обычно представляет из себя заранее собранные и оцифрованные тексты, которые сохраняются и используются для исследовательских целей. Статические корпусы могут использоваться для анализа языка, проведения исследований или обучения моделей машинного обучения.

3. Параллельный корпус - это набор текстов на двух или более языках, который является переводом друг друга. Параллельные корпусы используются для обучения и тестирования систем машинного перевода, а также для исследований сопоставительной лингвистики. Параллельные корпусы могут быть собраны с помощью текстов, которые имеют соответствующий перевод, или созданы специально для этой цели путем перевода существующих текстов.

4) Национальный корпус русского языка (НКРЯ) - это крупнейший лингвистический корпус русского языка, содержащий более 600 миллионов словоупотреблений. Он является национальным ресурсом, предназначенным для изучения русского языка в его различных аспектах.

НКРЯ включает в себя разнообразные тексты - художественные произведения, научные статьи, газетные и журнальные публикации, разговорную речь и многое другое. Это позволяет рассматривать различные виды языковых единиц, а также изучать их употребление в различных контекстах.

Кроме того, НКРЯ содержит различные инструменты и возможности для лингвистического исследования, такие как поиск по словоформам, лемматизация текстов, анализ частотности употребления слов и многое другое.

Таким образом, Национальный корпус русского языка является важным исследовательским ресурсом, который предоставляет ученым и лингвистам возможность изучать различные языковые явления и процессы на основе большого объема текстовых данных.

5) 1. Конкорданс - это представление слова или фразы в контексте текста, представленного в виде строки или небольшого текста. Конкорданс обычно используется в лингвистике и литературных исследованиях для исследования употребления слова или фразы.

2. Рандомизация - это процедура или метод, используемый в статистике, для случайного распределения участников или объектов исследования в разные группы или условия. Рандомизация помогает сделать результаты исследования более объективными и предотвратить искажение результатов.

3. Коллокация - это связь между словами, которая проявляется в их частом соседстве в тексте. Коллокации помогают понять, какие слова часто используются вместе и как они сочетаются друг с другом в естественном языке.

4. Подмассив - это часть массива данных, состоящая из определенного количества элементов, выбранных из исходного массива. Подмассивы используются для упрощения обработки больших объемов данных и выполнения операций над ними.

5. Парсинг - это процесс анализа текста или данных для извлечения определенной информации или структуры. Парсинг может использоваться для разбора синтаксиса предложений, обработки кода программ или извлечения данных из структурированных документов.

6. Лемматизация - это процесс приведения слова к его базовой (словарной) форме, называемой леммой. Лемматизация помогает унифицировать словоформы и упростить анализ текста, учитывая его смысловое значение.

7. Корпус-менеджер - это программное обеспечение, предназначенное для организации, управления и анализа корпусов текстов. Корпус-менеджеры позволяют исследователям лингвистам и специалистам по обработке текста работать с большими объемами данных и проводить различные исследования в области языкознания.

6) 1. The Corpus of Contemporary American English (COCA) - это крупнейший корпус английского языка, содержащий тексты с 1990 года по настоящее время. COCA позволяет исследователям изучать современное американское английское и проводить различные лингвистические исследования.

2. British National Corpus (BNC) - это корпус британского английского языка, содержащий разнообразные тексты с 20-го века. BNC является важным ресурсом для исследований по фонетике, грамматике и семантике английского языка.

Лабораторная работа 10

1)

|  |  |  |  |
| --- | --- | --- | --- |
|  | НКРЯ | COCA | BNC |
| Кол-во вхождений | 39076 / 315677 | 49260 | 5127 |

2)

|  |  |  |  |
| --- | --- | --- | --- |
| пример | НКРЯ | COCA | BNC |
| 1 | *Один из них спросил меня: " Хач? " Я ответил, что русский. Тогда они начали придираться к моему внешнему виду.* *(Анвар Тавобов. Террор несовершеннолетних // «Вечерняя Москва», 2002.02.07)* | " Alexander Solzhenitsyn # You must not have read Alexander Solzhenitsyn's analysis of **Russian** Socialism that is an above your comment in this thread. Here is Igor Shafarevich | The first-to-ten world title battle between the maverick American and the **Russian** is now all square after four games in Przno, Yugoslavia. |
| 2 | *Из слов, в разделении 52 ею означенных, русский человек ничего иного заключить не может, как только, что господин А. прав и что госпожа Всякая всячина его критиковала криво.* *(Н. И. Новиков. [Полемика с Екатериной II] (1769))* | On the issue of political asylum for immigrants, 200 **Russian** prisoners on U.S. soil in Fort Dix, New jersey had been captured in German | But Pallister found the nerve-racking task too much in the pouring rain of the **Russian** capital. |
| 3 | *Эдакое слово в русский язык тяжело вводить, надобно ввозить.* *(А. П. Сумароков. Чудовищи (1750))* | how Oil for Food ended up filling the pockets of French, German, and **Russian** middlemen and government agents, not to mention U.N. officials, and to understand | A **RUSSIAN** weightlifter was banned for life by the sport's governing body after rejecting an Olympic |

3)

|  |  |  |  |
| --- | --- | --- | --- |
| пример | НКРЯ | COCA | BNC |
| 1 | прилагательное | прилагательное | прилагательное |
| 2 | прилагательное | прилагательное | прилагательное |
| 3 | прилагательное | прилагательное | прилагательное |

4)

|  |  |  |  |
| --- | --- | --- | --- |
| пример | НКРЯ | COCA | BNC |
| 1 | Из России | Относится к стране | Из России |
| 2 | Из России | Из России | Относится к стране |
| 3 | язык | Относится к стране | Из России |

5) С морфологической точки зрения все слова являются прилагательными, а с точки зрения лексики имеют разные значения.

6) Рассмотренные корпусы могут быть использованы в лингвистическом исследовании для анализа частотности употребления определенных слов и выражений, изучения различий в употреблении языка в разных контекстах (например, в различных жанрах текстов или в разных временных периодах), исследования семантики слов и выражений, исследования синтаксиса и грамматики языка, а также для изучения стилистических особенностей текстов.