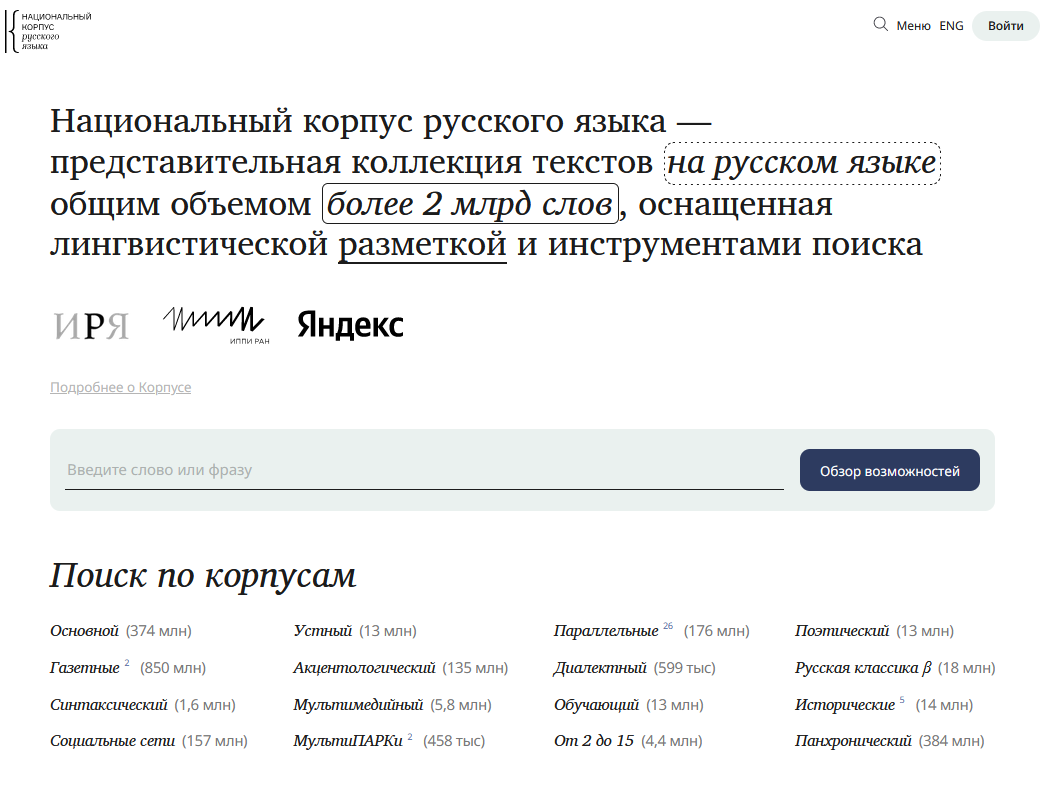
**Задание 1.**

**Национальный корпус русского языка**.

Ссылка на реусурс: <https://ruscorpora.ru/>



Информация о возможностях:

1. На сайте представлены корпуса(собрания текстов), используемые для решения лингвистических задач.
2. Корпуса НКРЯ можно применять для обучения нейросетей, анализа частоты появления слов в определенный период или синонимы к нему в тот же период.
3. Для использования надо ввести фразу/слово и из предложенных возможностей выбрать нужные.
4. В запрос подаем слово или фразу которую надо найти в корпусах НКРЯ
5. Преимущества:   
   НКРЯ представляет собой обширнейший ресурс аутентичных русских текстов разных стилей и форматов. Это позволяет создавать пользовательские корпуса в соответствии с целями обучения и интересами обучающихся.  
   Корпус также является незаменимым средством самопроверки студента и поддержкой самостоятельной работы с тем или иным аспектом языка.

С помощью НКРЯ преподаватель может осуществлять обучение вне зависимости от территориального расположения ученика, а также постепенно отказаться от печатного учебника как основного источника информации.

1. Недостатки:  
   Ограниченный объем: несмотря на обширность базы данных НКРЯ, она все же не охватывает все аспекты русского языка, что может быть недостатком при проведении специализированных исследований.  
   Для эффективного использования Национального корпуса русского языка на уроках необходимо обладать определенными навыками работы с компьютером и поисковыми системами.
2. Похожие ресурсы:  
   <http://corpus.byu.edu/> - Коллекция различных лингвистических корпусов на английском языке, собранная профессором лингвистики Марком Девисом.   
   <http://bnkorpus.info/> - Белорусский корпус
3. Исследования:   
   *Савчук, С. О.*Корпус региональных газет России и зарубежья // Труды Института русского языка им. В.В. Виноградова. Вып. 6. М., 2015. C. 163–193.

*Савчук, С. О.* Корпус как инструмент для исследования особенностей функционирования русского языка в региональной прессе // Труды Института русского языка им. В.В. Виноградова. Вып. 6. М., 2015. C. 333-364. Савчук С.О. Лексико-семантические особенности текстов региональных СМИ по данным корпусного обследования // Труды международной конференции «Корпусная лингвистика – 2015». СПб, 2015. С. 398–407.

1. Пример своего запроса: Вводим слово “новинка”, нам выдает точное количество слова в корпусах, историю употребления словоформ, грамматику, семантику. Таким образом мы можем проанализировать словоупотребления и в целом получить информацию о словах, которая необходима для работы.

**E-library**

Ссылка на ресурс: <https://elibrary.ru/defaultx.asp>

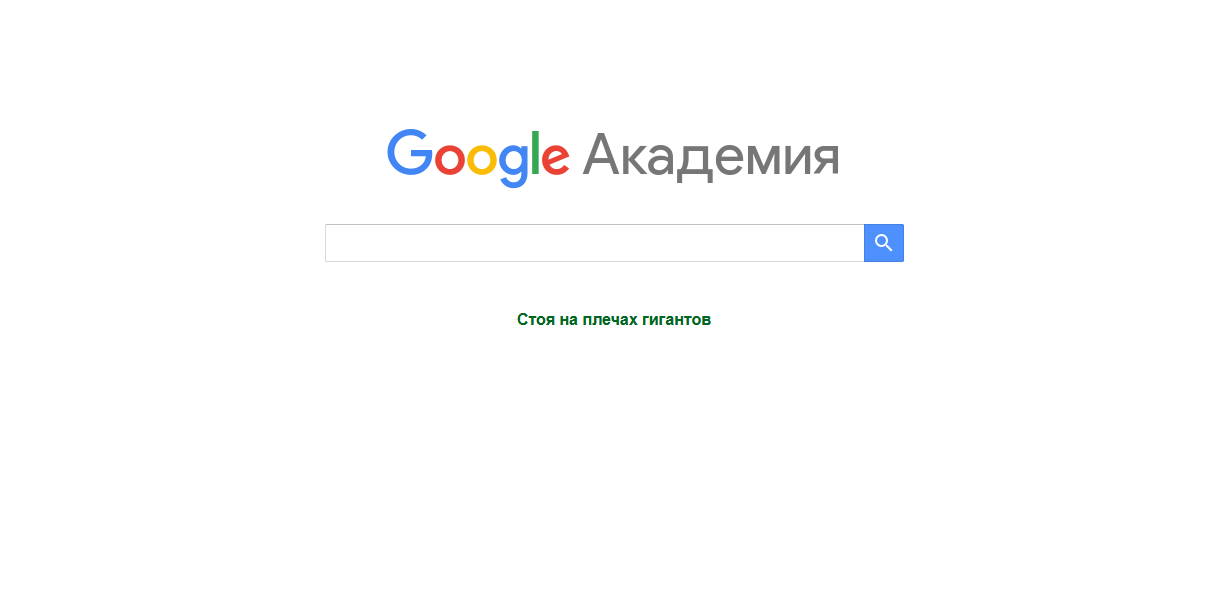


Информация о ресурсе:

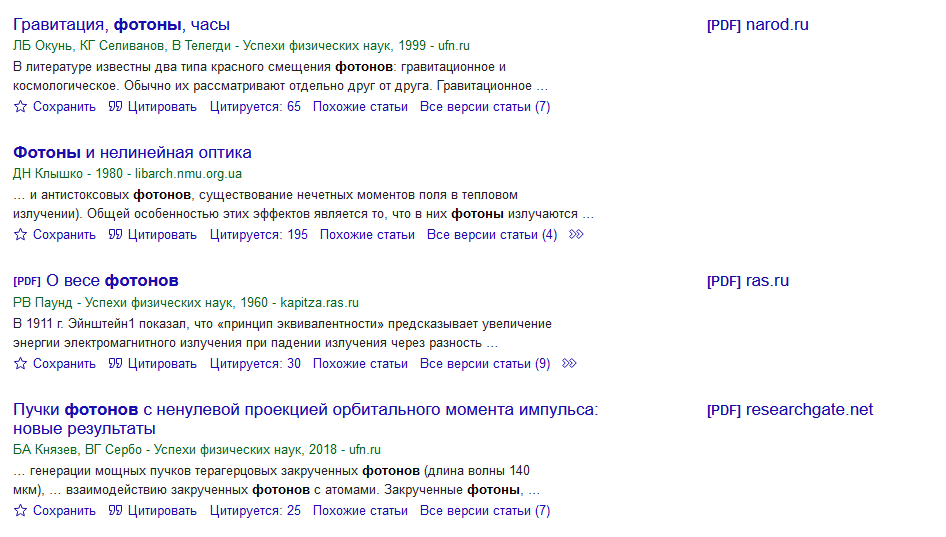
1. Множество статей на разнообразные темы
2. На статьи можно опираться при исследованиях.
3. Вводим запрос в поисковое окно, анализируем результаты
4. В запрос подаем название статьи или темы, которую надо найти в электронной библиотеке
5. Преимущества:  
   множество статей на разные темы  
   большой доступный бесплатный функционал
6. Недостатки:  
   Присутствуют устаревшие, псевдонаучные статьи  
   Неудобный интерфейс
7. Похожие ресурсы:  
   <https://scholar.google.ru/schhp?hl=ru>  
   <https://cyberleninka.ru/>
8. Исследования:  
   <https://cyberleninka.ru/article/n/analiz-publikatsionnoy-aktivnosti-i-vklyuchennosti-v-mezhdunarodnye-nauchnye-kommunikatsii-s-pomoschyu-bibliometricheskih?ysclid=ltmm5ndanp885883399>
9. Пример взаимодействия: вводим слово “корпус” и получаем множество статей на эту тему или связанную с ней. Как результат: получаем множество информации на эту тематику.

**Google Академия**

Ссылка на ресурс: <https://scholar.google.ru/schhp?hl=ru>

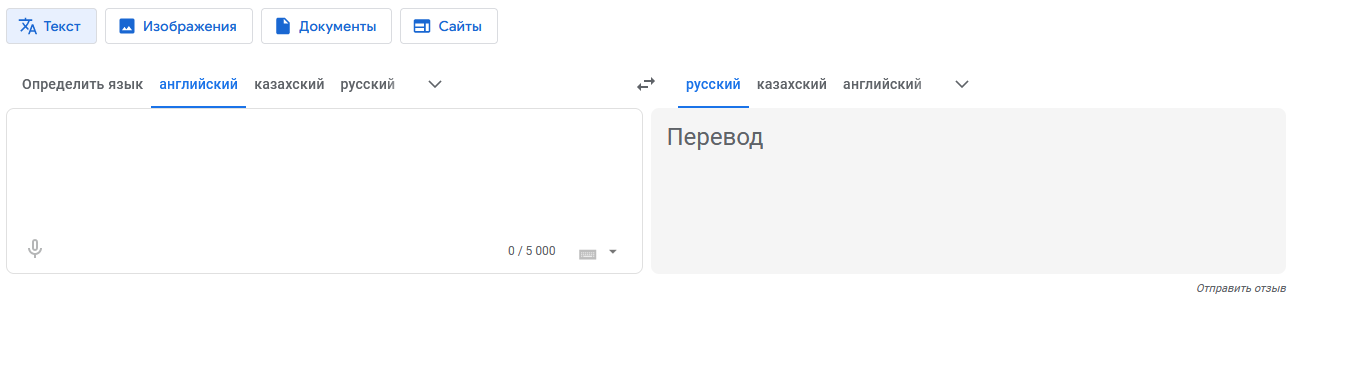


Информация о ресурсе:

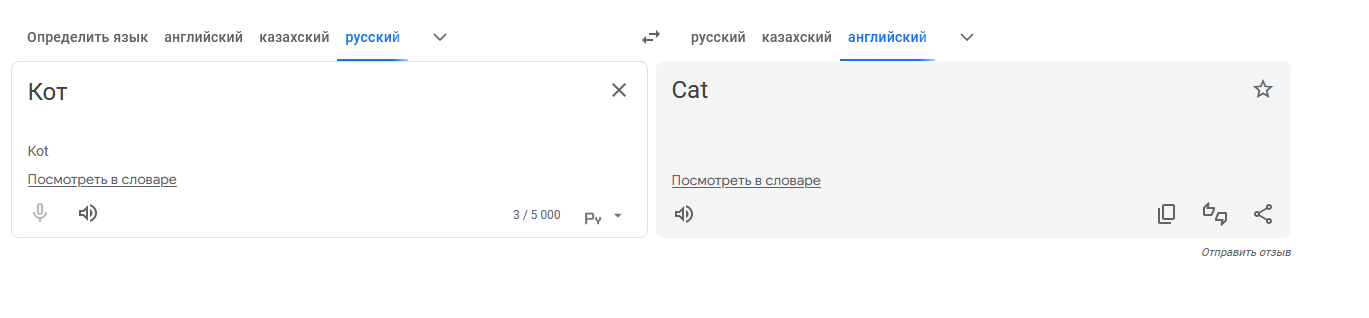
1. Ресурс предоставляет доступ к поиску среди всех статей в среде Интернет.
2. Полученную информацию можно применять в исследовании
3. Вводим запрос в поисковое окно, анализируем результаты
4. В запрос подаем название статьи или темы, которую надо найти в поиске
5. Преимущества:  
   Легкий и быстрый доступ к поисковой системе. Если необходимо быстро найти материалы, необязательно регистрироваться на сайте и не нужно вносить плату за подписку.  
   Удобный поиск по различным критериям – по автору, тематике, дате и месту публикации, по отдельным изданиям и хранилищам.  
   Обширный охват проиндексированных научных трудов, собранных по всему миру, как периодических изданий, так и непериодических сборников, учебных пособий и других работ, представленных в открытом доступе в интернете.
6. Недостатки:  
   Недостаточный охват отдельных изданий, которые отказались включить индексацию их журналов, что может стать серьезным препятствием для некоторых ученых.  
   Из-за стремления Google Scholar добиться максимальной индексации самых разных научных трудов в поисковой базе могут попадаться низкокачественные и псевдонаучные публикации.
7. Похожие ресурсы:  
   <https://cyberleninka.ru/>  
   <https://elibrary.ru/defaultx.asp>
8. Исследования:  
   Бизенков Е.А. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ПОИСКОВОЙ И НАУКОМЕТРИЧЕСКОЙ ПЛАТФОРМЫ GOOGLE SCHOLAR (АКАДЕМИЯ GOOGLE) // Международный журнал прикладных и фундаментальных исследований. – 2017. – № 10-1.
9. Пример использования: Вводим запрос “фотоны” и на выходе получаем набор статей. Проанализировав выданный результат приходим к выводу, что система выдает научные статьи полезные при исследованиях фотонов

**Google Translate**

Ссылка на ресурс: <https://translate.google.com/>



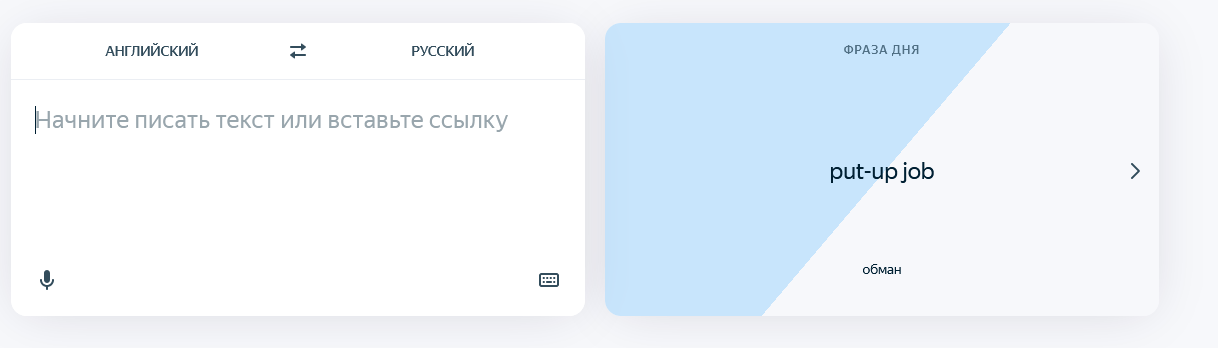
Информация о ресурсе:

1. Позволяет переводить слова с одного языка на другой.
2. Можно использовать для перевода неизвестного языка, общения, исследований в лингвистике.
3. Вводим слово/фразу/отрывок текста в соответствующее окошко
4. Стараемся грамотно ввести переводимый фрагмент
5. Преимущества:  
   Быстрая скорость перевода  
   Не требует регистрации  
   Высокое качество перевода отдельных слов/фраз
6. Недостатки:  
   Не хранит контекст  
   Есть проблемы с переводом больших текстов
7. Похожие ресурсы:  
   <https://translate.yandex.ru/>  
   <http://www.promt.ru/>
8. Исследования:  
   <https://eee-science.ru/item-work/2020-2409/?ysclid=ltmokzws9o307335069>  
   <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-onlayn-perevodchikov?ysclid=ltmpvh596e509953736>
9. Пример использования:

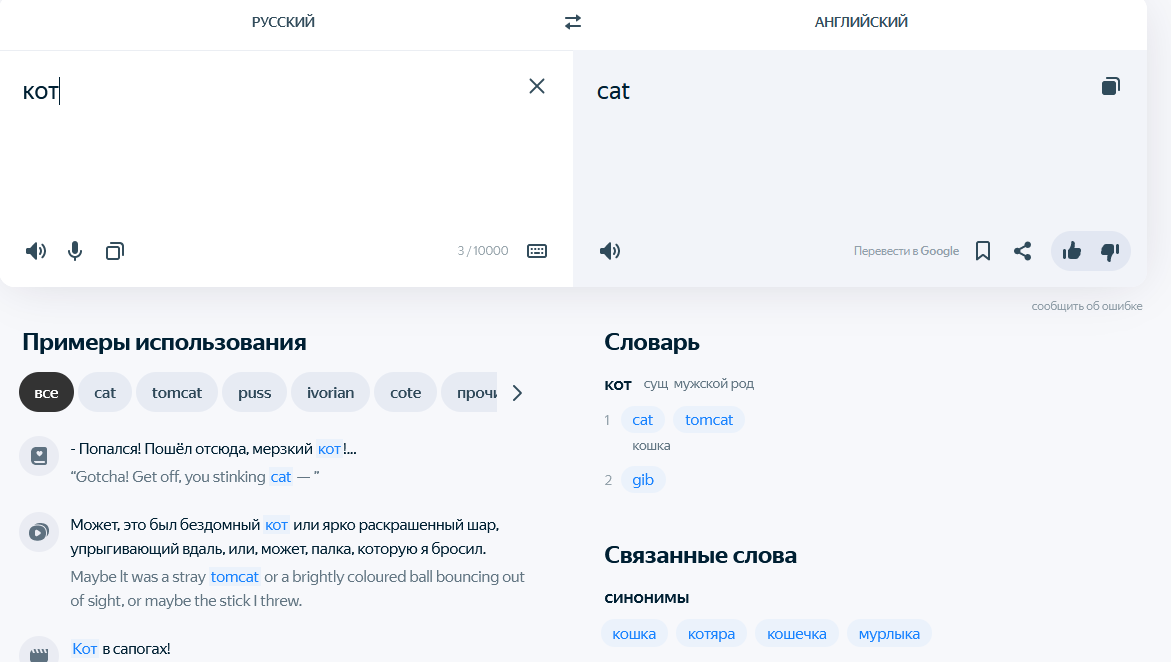
Вводим слово “кот”, получаем перевод, без контекста.

**Яндекс переводчик**

Ссылка на ресурс: <https://translate.yandex.ru/>



Информация о ресурсе:

1. Позволяет переводить слова с одного языка на другой.
2. Можно использовать для перевода неизвестного языка, общения, исследований в лингвистике.
3. Вводим слово/фразу/отрывок текста в соответствующее окошко
4. Стараемся грамотно ввести переводимый фрагмент
5. Преимущества:  
   Быстрая скорость перевода  
   Не требует регистрации  
   Высокое качество перевода отдельных слов/фраз
6. Недостатки:  
   Не хранит контекст  
   Есть проблемы с переводом больших текстов
7. Похожие ресурсы:  
   <https://translate.google.com/>  
   <http://www.promt.ru/>
8. Исследования:  
   <https://eee-science.ru/item-work/2020-2409/?ysclid=ltmokzws9o307335069>  
   <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-onlayn-perevodchikov?ysclid=ltmpvh596e509953736>
9. 

Вводим запрос “кот” - получаем перевод, связанные слова и примеры использования

**ЗАДАНИЕ 2**

Вопросы для обсуждения:

1. Единицей корпуса являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной проблемной области.
2. Корпусы составляют в зависимости от поставленных целей и задач, они должны отражать текстовое разнообразие(если это необходимо), иметь специализированность или общую направленность текстов.
3. Исследовательскими называются такие корпусы, которые предназначены преимущественно для изучения различных аспектов функционирования языковой системы. Они строятся не post factum — после проведения какого-либо исследования, а до его проведения.   
   Статический корпус – корпус, отражающий определенное временное состояние языковой системы.   
   Параллельный корпус — это двуязычный корпус, то есть текст оригинала и его перевод на какой-то другой язык, причем эти два текста не просто лежат рядом друг с другом, а должны быть выровнены: отдельные фрагменты оригинала должны совпадать с соответствующими фрагментами перевода.

| **Критерии классификации** | **Виды корпусов** |
| --- | --- |
| Тип языковых данных | Устные / Письменные / Смешанные |
| Формат представления данных | Мономодальные / Мультимодальные (мультимедийные) |
| Объем и полнота данных | Общие (представительные, национальные) / Для специальных целей: мониторинговые, иллюстративные |
| Число представленных языков | Одноязычные / Двуязычные / Многоязычные |
| Число языков в одном тексте (для дву- и многоязычных корпусов) | Параллельные / Смешанные |
| Выравнивание текстов на разных языках (для дву- и многоязычных корпусов) | Соизмеримые или параллельные / Несоизмеримые или непараллельные |
| Наличие единых принципов отбора языкового материала | Сопоставимые / Несопоставимые |
| Представленность языкового материала | Полнотекстовый / Фрагментированный (в т.ч. корпуса n-грамм и конкордансы) |
| Временной период, к которому относятся собранные языковые данные | Синхронные / Диахронные |
| Наличие разметки | Размеченные (аннотированные) / Неразмеченные (неаннотированные) |
| Тип разметки (для размеченных корпусов) | Метатекстовая разметка / Лингвистическая разметка (морфологическая, синтаксическая, семантическая) / Экстралингвистическая разметка |
| Доступность | Свободно доступные / Частично доступные / Закрытые (коммерческие) |

1. Британский национальный корпус: содержит около 100 миллионов словоупотреблений, собранных из различных текстов на английском языке.  
   По виду корпусов:   
   По типу текстов: В его состав входят тексты разных жанров и стилей.  
   По разнообразию языков: Корпус включает в себя множество форм английского языка, что можно использовать для исследований(Многоязычные)  
   По временному периоду: представлены тексты разных периодов(Диахронные)
2. Глоссарий по теме «Корпусная лингвистика» :  
   ***Конкорданс*** - алфавитный перечень всех слов какого-либо текста с указанием контекстов их употребления; тип словаря, представленный в таком виде  
   ***Рандомизация*** - процедура случайного выбора элементов статистической совокупности при проведении выборочного исследования  
   ***Коллокация*** - в корпусной лингвистике под коллокацией понимается последовательность слов или терминов, частотность совместного появления которых в корпусе выше, чем ожидаемая вероятность их совместного появления.  
   ***Подмассив*** - это любая последовательность чисел в массиве (непрырывная).  
   ***Парсинг*** - это синтаксический анализ, который автоматически производится парсером – специальной программой или скриптом.  
   ***Лемматизация*** - процесс привода словоформы к лемме — её нормальной (словарной) форме.  
   ***Корпус- менеджер* -** программное обеспечение для корпуса**,** котороеподсчитывает конкретные словоформы, группы словосочетаний и также может выводить результаты статистики, представить в виде диаграммы.
3. Среди обнаруженных мною ресурсов интерес представляют 2:  
   первый - <http://blogs.it-claim.ru/izelentsov/>, где представленны 4 корпуса, созданные группами специалистов.   
   второй - <https://elementy.ru/catalog?type=127&ysclid=ltn8egrg4e692471229>, где собрана информация о корпусах

**Лабораторная работа 10**

1. Результаты представленные в таблице могут свидетельствовать о том, что наименьшее количество вхождений ключевого слова в Британском корпусе

|  | НКРЯ | COCA | BNC |
| --- | --- | --- | --- |
| Количество вхождений | 39 076(315 677) | 49303 | 5370 |

| № Примера | НКРЯ | COCA | BNC |
| --- | --- | --- | --- |
| 1 | Я не люблю *русский* рок, но эта одна из тех 5-6 групп, которые я слушаю [Переписка в icq между agd-ardin и Колючий друг (08.02.2008)] | It's just very expensive. # As for the mirror experiment, the **Russian** version failed rather quickly because they used a robotic lander, and it was unableх[Moon hoax: why not use telescopes to look at the landers? | Bad] | Here is a description by Diderot of a [**Russian**](http://bncweb.lancs.ac.uk/cgi-binbncXML/context.pl?thin=0&numOfFiles=859&inst=50&qname=dercomputer_1710219560&max=108&numOfSolutions=5370&queryID=dercomputer_1710219560&thMode=M5370%23859%23no_subcorpus%23%23&theID=dercomputer_1710219560&theData=%5Bword%3D%22Russian%22%25c%5D&chunk=1&queryMode=simple&view2=nonrandom&simpleQuery=Russian&program=search&view=list&listFiles=0&phon=0&subcorpus=no_subcorpus&queryType=CQL&qtype=0&text=A04&refnum=2&theShowData=Russian&len=-24&showTheTag=0&color=0&begin=1358&token_offset=8&nodeCount=1&hitSunit=1358&spids=1&interval=11&urlTest=yes) country scene by Le Prince, from criticism of the Salon in 1767. [Art criticism: a user's guide.] |
| 2 | Мое сознание постепеннно забыло что *русский* — это мой язык " by default " [Запись LiveJournal (2004)] | Now what we we talking about? # Although you can say that the **Russian** mirror performance is poorer than the US ones in some respects, as described above[Moon hoax: why not use telescopes to look at the landers? | Bad] | The shrugger doesn't care whether he lives or dies — and the designer duel arranged for himself by this divided and indifferent man is a form of [**Russian**](http://bncweb.lancs.ac.uk/cgi-binbncXML/context.pl?thin=0&numOfFiles=859&inst=50&qname=dercomputer_1710219560&max=108&numOfSolutions=5370&queryID=dercomputer_1710219560&thMode=M5370%23859%23no_subcorpus%23%23&theID=dercomputer_1710219560&theData=%5Bword%3D%22Russian%22%25c%5D&chunk=1&queryMode=simple&view2=nonrandom&simpleQuery=Russian&program=search&view=list&listFiles=0&phon=0&subcorpus=no_subcorpus&queryType=CQL&qtype=0&text=A05&refnum=3&theShowData=Russian&len=-32&showTheTag=0&color=0&begin=933&token_offset=28&nodeCount=1&hitSunit=933&spids=1&interval=11&urlTest=yes) roulette. [Authors] |
| 3 | От самого легкого — к самому трудному: английский, немецкий, *русский*, французский (с транслита) [Наши дети: Подростки (2004)] | mith &; Nephew countered with its own experts, including a top deputy of the **Russian** Federation's Department of Libraries, flown into Houston to be deposed. # KCI[The bloody patent battle over a healing machine - Fortune] | Even so, he wrote, what might, à la rigueur , be acceptable for a [**Russian**](http://bncweb.lancs.ac.uk/cgi-binbncXML/context.pl?thin=0&numOfFiles=859&inst=50&qname=dercomputer_1710219560&max=108&numOfSolutions=5370&queryID=dercomputer_1710219560&thMode=M5370%23859%23no_subcorpus%23%23&theID=dercomputer_1710219560&theData=%5Bword%3D%22Russian%22%25c%5D&chunk=1&queryMode=simple&view2=nonrandom&simpleQuery=Russian&program=search&view=list&listFiles=0&phon=0&subcorpus=no_subcorpus&queryType=CQL&qtype=0&text=A08&refnum=6&theShowData=Russian&len=-56&showTheTag=0&color=0&begin=512&token_offset=17&nodeCount=1&hitSunit=512&spids=1&interval=11&urlTest=yes) Grandmaster seems merely perverse in a retired artist. [The big glass] |

| № примера | НКРЯ | COCA | BNC |
| --- | --- | --- | --- |
| 1 | прилагательное | прилагательное | прилагательное |
| 2 | прилагательное | прилагательное | прилагательное |
| 3 | прилагательное | прилагательное | прилагательное |

| № примера | НКРЯ | COCA | BNC |
| --- | --- | --- | --- |
| 1 | из россии | российская версия | принадлежность к россии |
| 2 | язык | российская версия | устойчивое выражение |
| 3 | язык | название страны | принадлежность к стране |

1. В моем случае все взятые слова с точки зрения морфологии являются - прилагательными, а с лексической точки зрения они различаются: где-то это отношение к стране, где-то выражение и т.д.
2. Рассмотренные корпуса можно использовать для сравнения частоты появления слова в текстах данных корпусов, сравнение их лексических значений и многое другое.