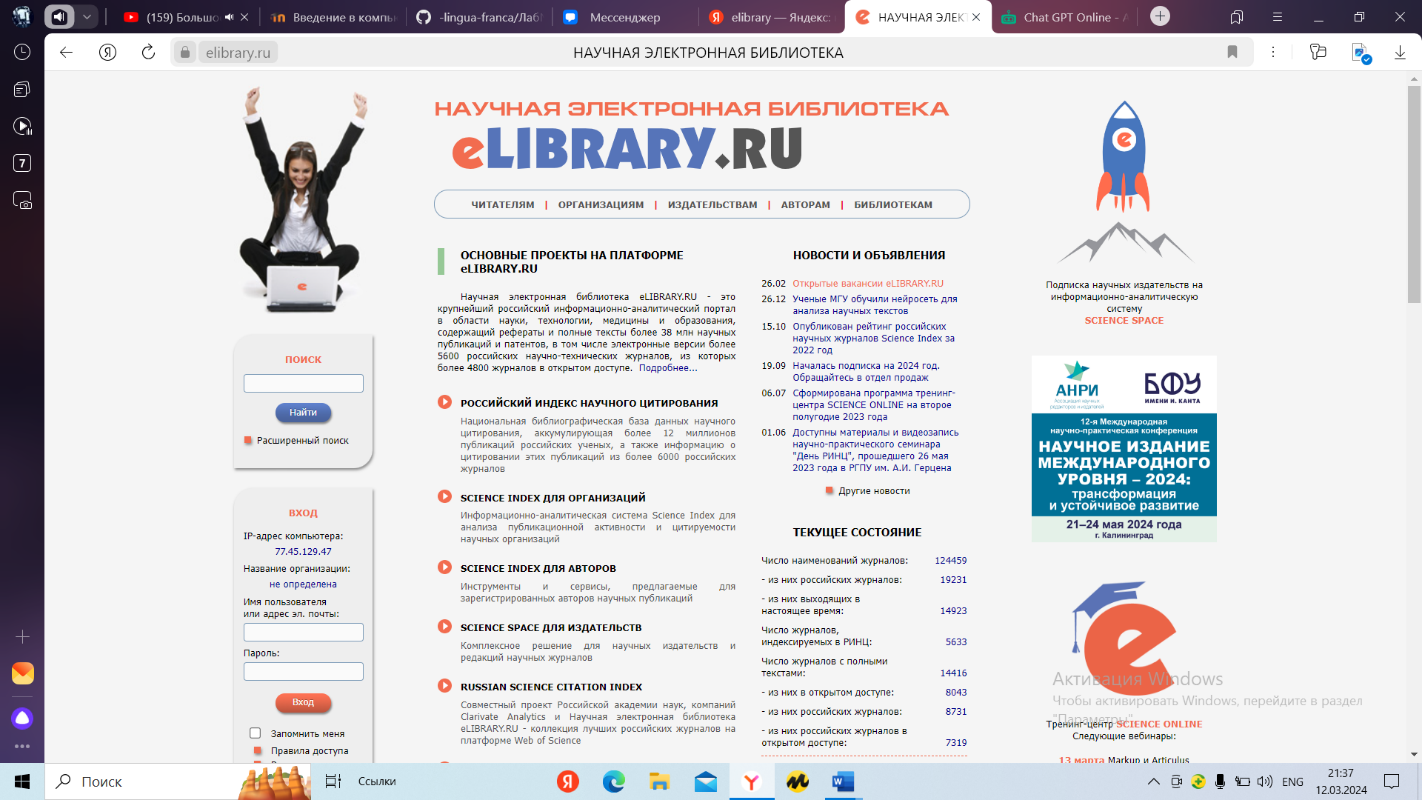
1. **ELIBRARY.RU (**[**https://elibrary.ru/defaultx.asp**](https://elibrary.ru/defaultx.asp)**?)**

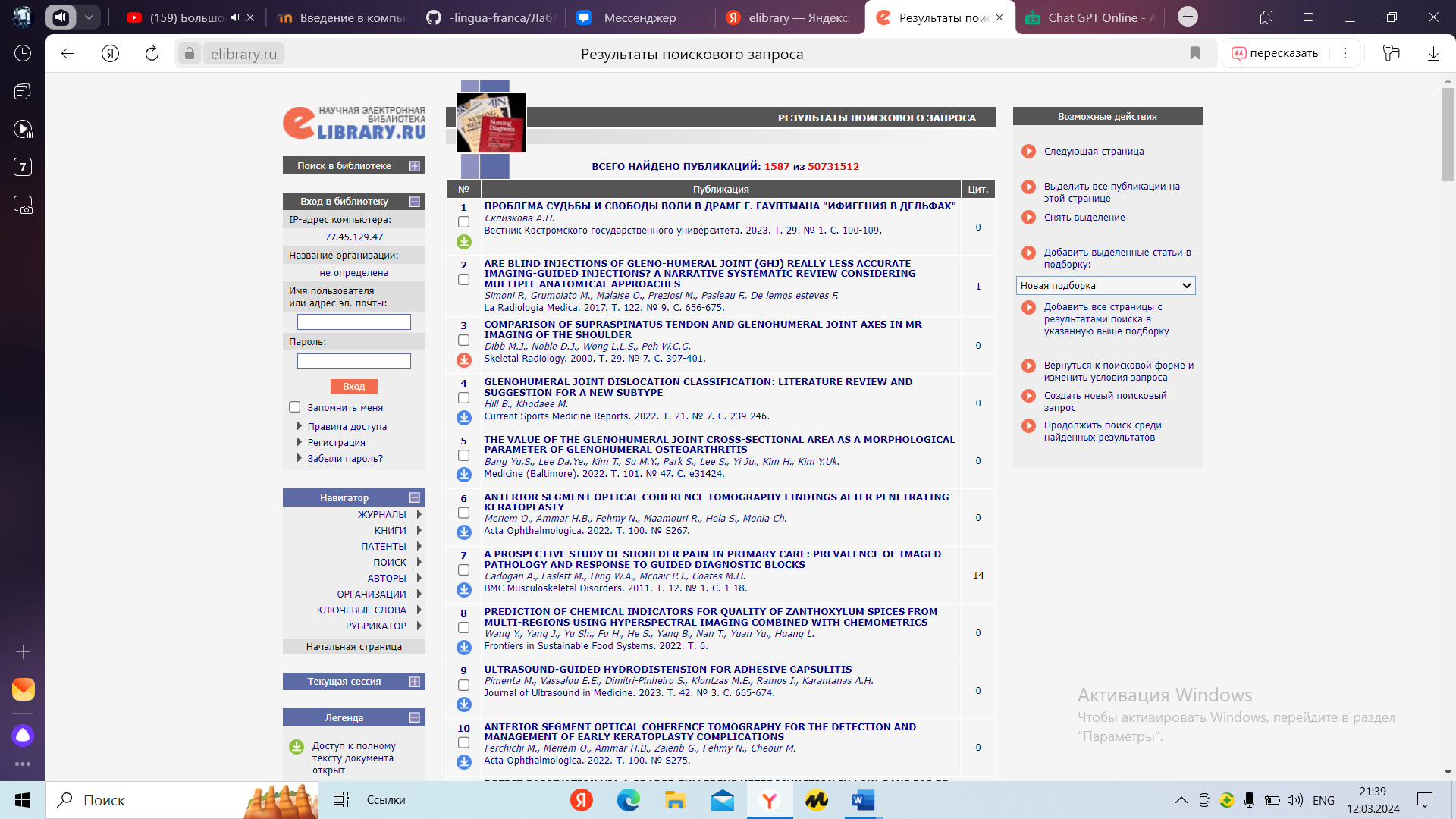
Крупнейший российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты более 22 млн научных статей и публикаций, в том числе электронные версии более 4800 российских научно-технических журналов, из которых более 3800 журналов в открытом доступе.

eLIBRARY.RU - крупнейшая в России электронная библиотека научных публикаций, обладающая богатыми возможностями поиска и получения информации. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ) - созданным по заказу Минобрнауки РФ бесплатным общедоступным инструментом измерения и анализа публикационной активности ученых и организаций. eLIBRARY.RU и РИНЦ разработаны и поддерживаются компанией "Научная электронная библиотека".

"The effects of mindfulness meditation on stress reduction in college students" - автор: Brown, S.

"The relationship between physical activity and mental health in elderly adults" - автор: Thompson, M.

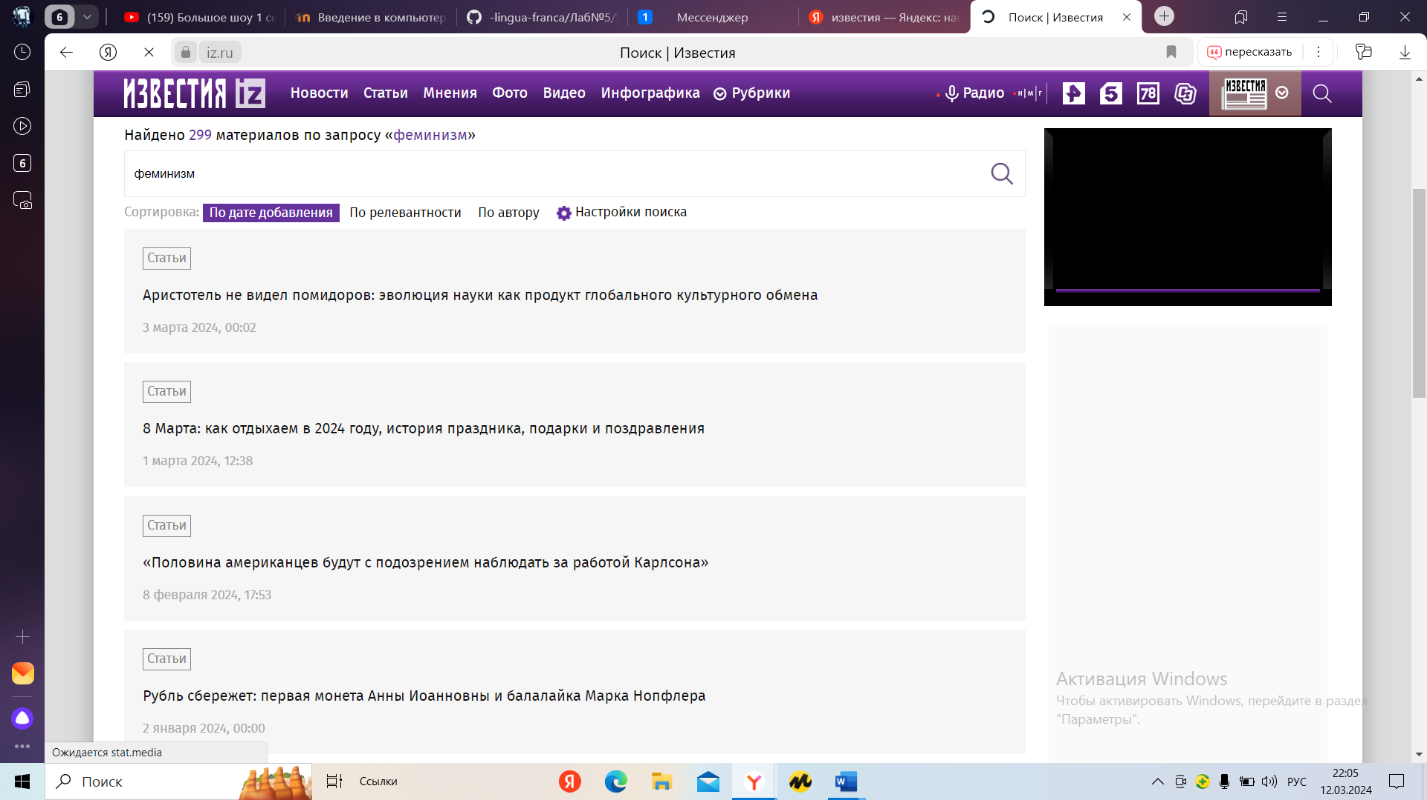
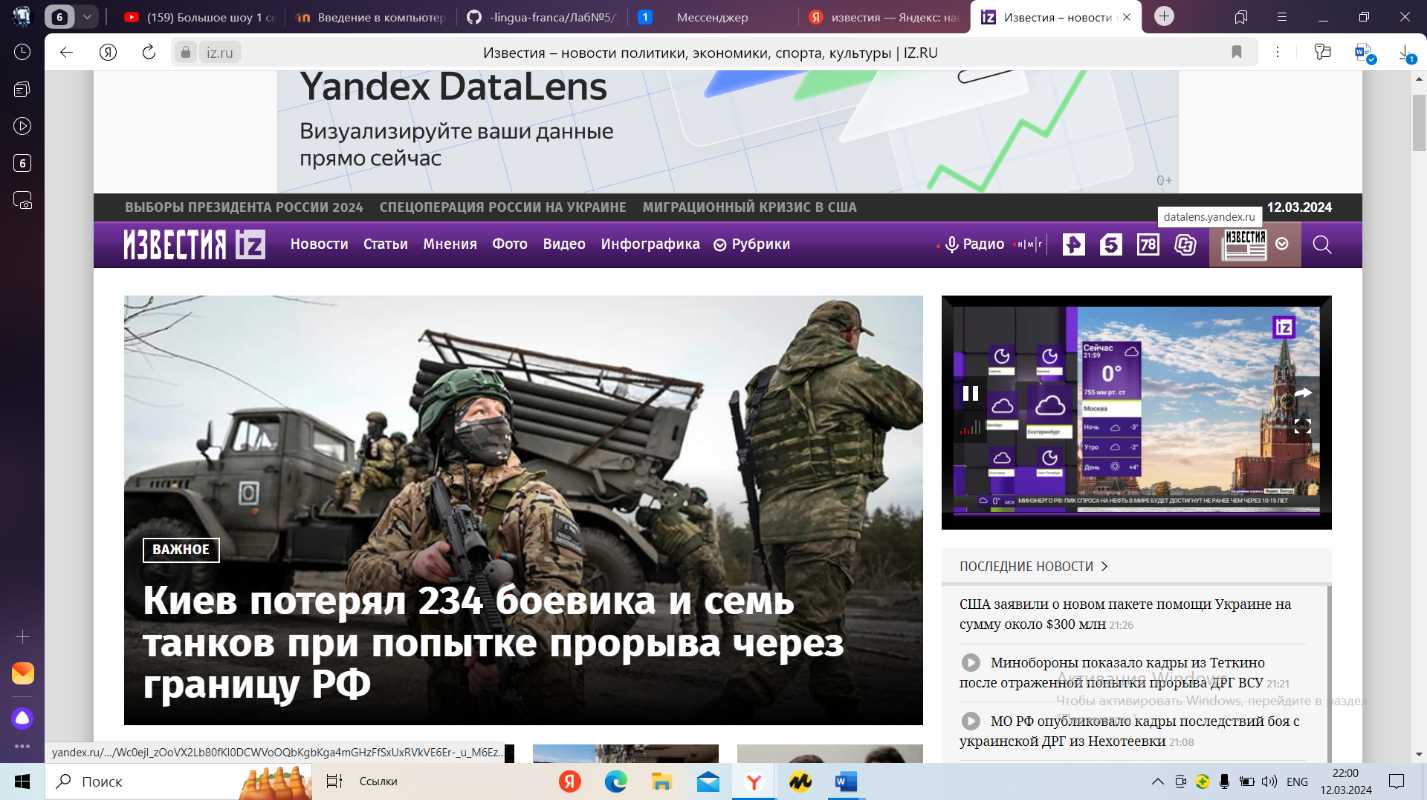


Чтобы произвести запрос существует окно поиск, куда вводятся название материалов и нажимаем кнопку найти. Мы переходим на страницу со всеми похожими названиями материалов и выбираем то, что нам нужно

1. Известия (<https://iz.ru/>)

Лента новостей, обновляемая в режиме реального времени публикациями о событиях политики, экономики, социальной жизни, культуры, спорта и др.

Этот ресурс можно использовать для получения необходимой информации о каких-либо событиях. В нем можно найти информацию о том, что происходит в мире, о достижениях спортсменов и тд. В нем содержаться фото, видео, статьи, новости, мнения экспертов и не только о чем-либо, всевозможные рубрики. Чтобы осуществить запрос нужно щелкнуть по значку лупы в правом верхнем углу и сделать запрос о той или иной теме, после чего вы попадаете на страницу, где вам представлены всевозможные статьи об этой теме

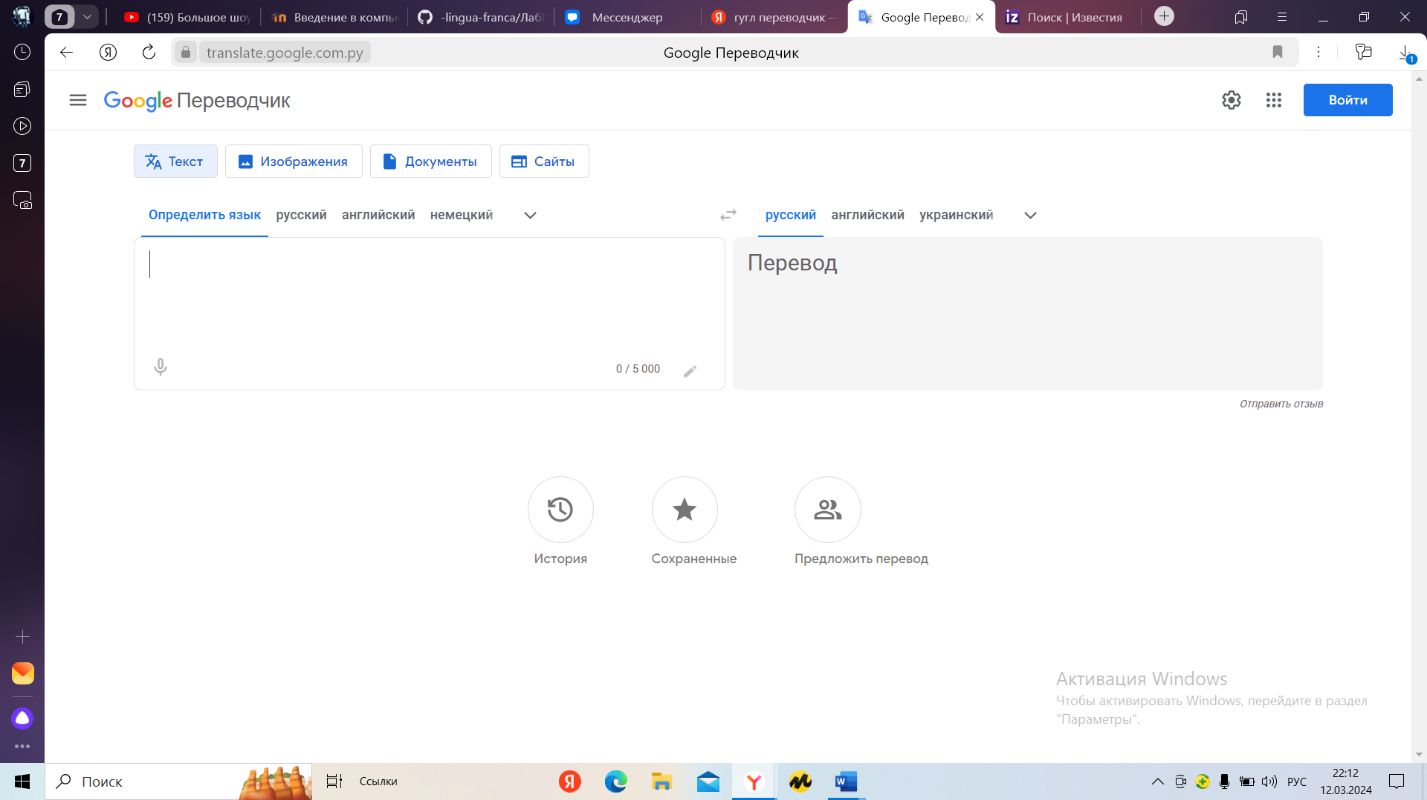


Преимущества этого ресурса состоят в том, что очень он очень прост в эксплуатации, можно всегда найти свежую информацию и очень давнюю

Недостатки заключаются в том, что дизайн устаревший

3. Google Translate (<https://translate.google.com.py/?hl=ru>)

* [**Google Translate**](https://translate.google.com/) Бесплатный веб-сервис компанииGoogle, предназначенный для перевода отдельных слов,текстов или целых веб-страниц. Преимущество этого переводчика заключается в том, что в отличии от остальных он работает на базе собственного автономного программного обеспечения. Google Переводчик имеет достаточно широкий спектр возможностей. С помощью него стал возможен перевод целых веб-страниц, а также одновременный поиск информации с параллельным переводом на любой другой язык. Сервис моментально осуществляет перевод текстов. Кроме этого возможен перевод на 103 языка мира.



Этот переводчик может переводить не только записанный текст, но и текст на фото сайты, документы, что упрощает пользователю работу. Чтобы перевести текст нужно записать его в левое окно, выбрав язык на котором он написан изначально, и выбрать язык на который нужно перевести. Для перевода текста на фото нужно загрузить это фото на сайт и выбрать язык на который собираетесь переводить.

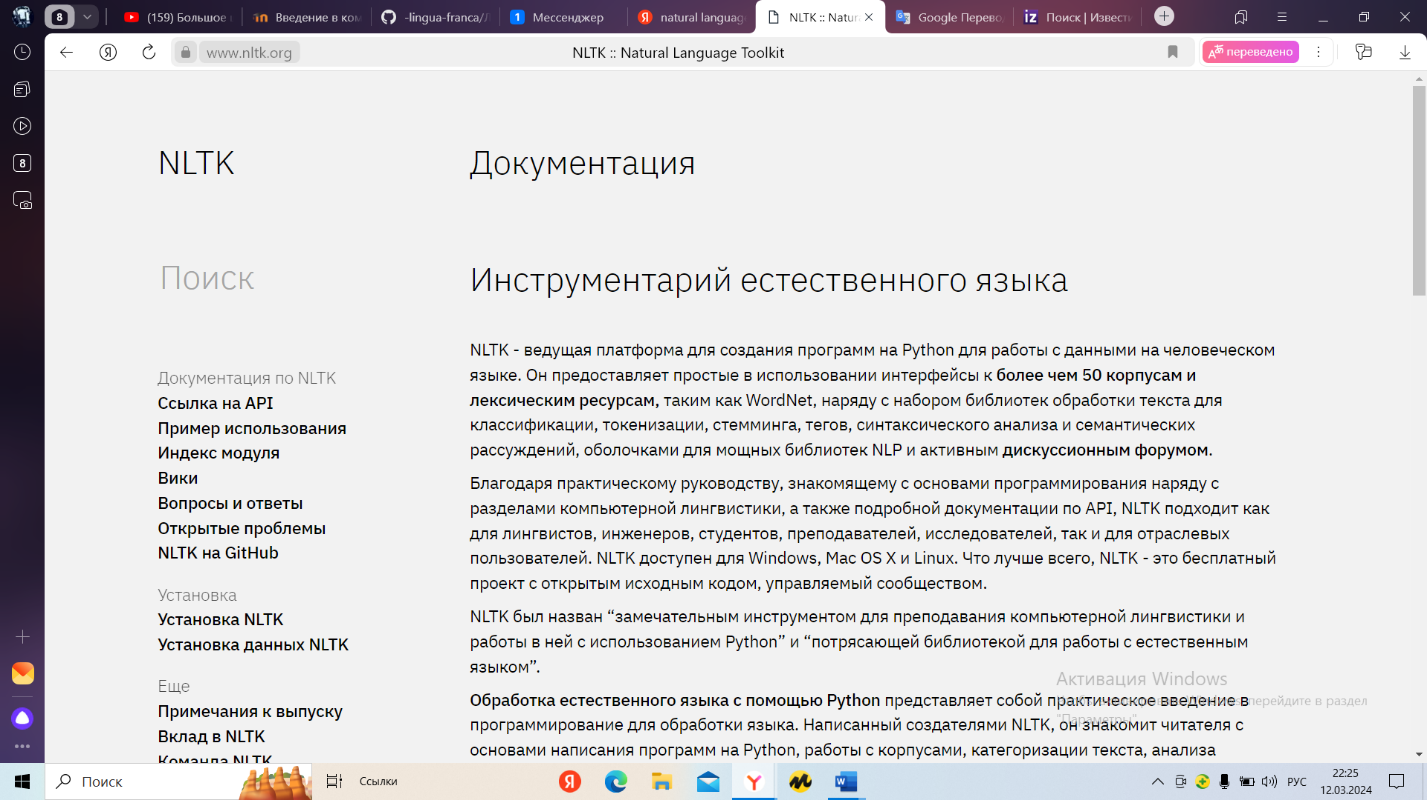
Достоинство этого ресурса заключается в том, что пользоваться им может даже ребенок. Простое оформление где нельзя запутаться. Множество разных языков. Можно даже прослушать как произноситься переведенный текст(не на всех языках).

Недостаток заключается в том, что перевод не всегда точный.

# 4. [NLTK](https://www.nltk.org/)

[**Natural Language Toolkit**](http://www.nltk.org/book/)Пакет [библиотек](https://ru.wikipedia.org/wiki/%D0%91%D0%B8%D0%B1%D0%BB%D0%B8%D0%BE%D1%82%D0%B5%D0%BA%D0%B0_%28%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5%29) и программ для символьной и статистической обработки естественного языка, написанных на языке программирования Python. Содержит графические представления и примеры данных. Сопровождается обширной документацией, включая книгу с объяснением основных концепций, стоящих за теми задачами обработки естественного языка, которые можно выполнять с помощью данного пакета.

NLTK хорошо подходит для студентов, изучающих компьютерную лингвистику или близкие предметы, такие как эмпирическая лингвистика, когнитивистика, искусственный интеллект, информационный поиск и машинное обучение.NTLK с успехом используется в качестве учебного пособия, в качестве инструмента индивидуального обучения и в качестве платформы для прототипирования и создания научно-исследовательских систем. NTLK является свободным программным обеспечением. Проект возглавляет Стивен Бёрд.



NLTK - ведущая платформа для создания программ на Python для работы с данными на человеческом языке. Он предоставляет простые в использовании интерфейсы к [более чем 50 корпусам и лексическим ресурсам,](https://www.nltk.org/nltk_data/) таким как WordNet, наряду с набором библиотек обработки текста для классификации, токенизации, стемминга, тегов, синтаксического анализа и семантических рассуждений, оболочками для мощных библиотек NLP и активным [дискуссионным форумом](https://groups.google.com/group/nltk-users).

Благодаря практическому руководству, знакомящему с основами программирования наряду с разделами компьютерной лингвистики, а также подробной документации по API, NLTK подходит как для лингвистов, инженеров, студентов, преподавателей, исследователей, так и для отраслевых пользователей. NLTK доступен для Windows, Mac OS X и Linux. Что лучше всего, NLTK - это бесплатный проект с открытым исходным кодом, управляемый сообществом.

NLTK был назван “замечательным инструментом для преподавания компьютерной лингвистики и работы в ней с использованием Python” и “потрясающей библиотекой для работы с естественным языком”.

[Обработка естественного языка с помощью Python](https://www.nltk.org/book/) представляет собой практическое введение в программирование для обработки языка. Написанный создателями NLTK, он знакомит читателя с основами написания программ на Python, работы с корпусами, категоризации текста, анализа лингвистической структуры и многим другим.

Чтобы осуществить поиск нужно нажать на серый текст ,,Поиск,, и в открывшемся окне записать то, что нам нужно

Достоинства: большая база информации

Недостатки: непонятный в использовании в начале

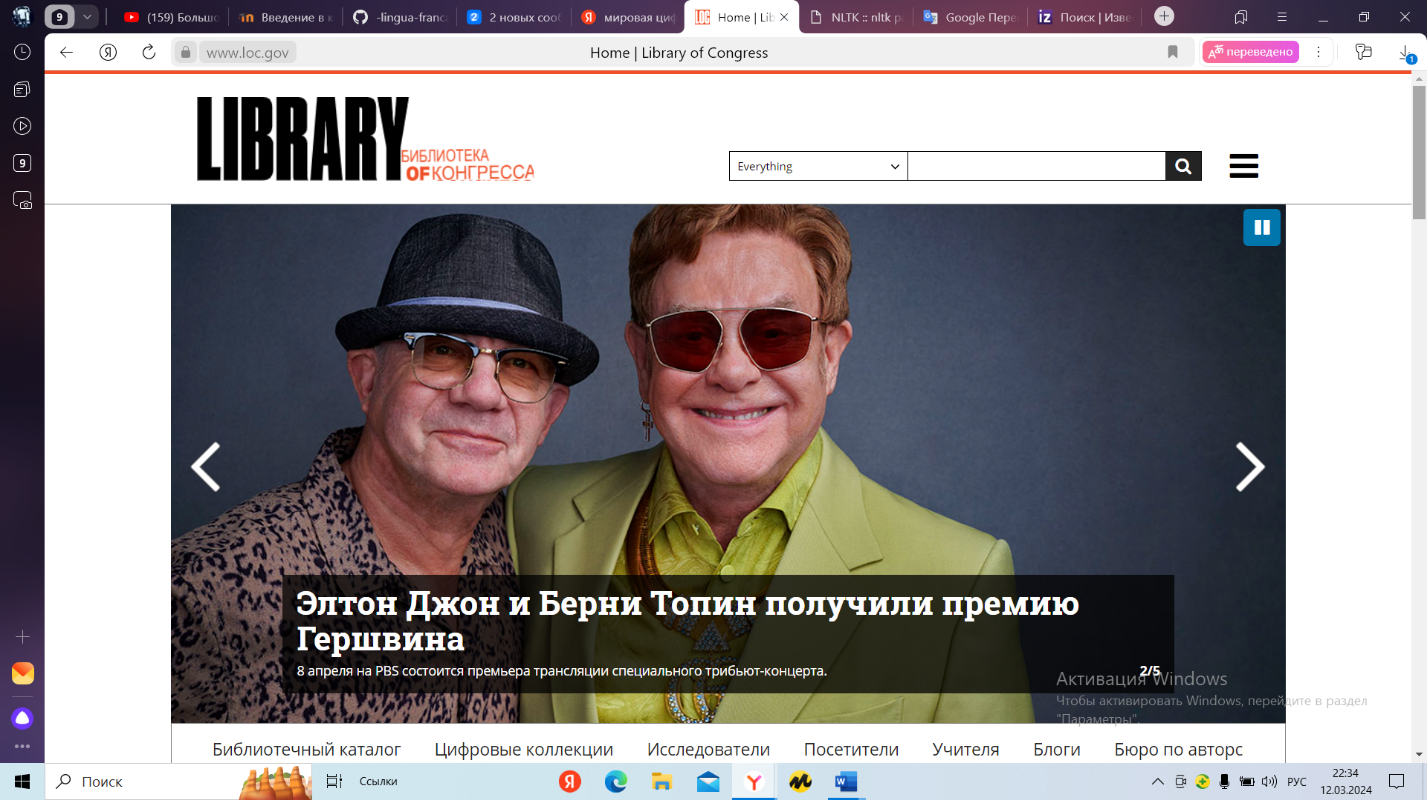
5. Мировая цифровая библиотека (<https://www.loc.gov/collections/world-digital-library/about-this-collection/>)

Мировая цифровая библиотека (англ. World Digital Library) — проект Библиотеки Конгресса США.

В библиотеке собраны оцифрованные версии ценнейших материалов по истории и культуре.

В 2007 году к проекту присоединилась Российская национальная библиотека. Участниками международного проекта являются национальные книгохранилища и архивы различных стран.

Мировая цифровая библиотека предоставляет бесплатный доступ в сети Интернет в многоязычном формате к большому количеству материалов, представляющих культуры разных стран мира.



Достоинства: Огромное количество информации на любые темы

Недостатки: Сайт не полностью на русском

Вопросы:

1.Единицей корпуса может быть отдельный текстовый документ, предложение, слово или даже символ, в зависимости от того, какую единицу анализа выбрал исследователь. Для текстового анализа наиболее распространенной единицей является предложение или документ, в то время как для лингвистического анализа часто используется слово или символ (например, фонема или морфема).

2. Отбор текстов для корпуса зависит от целей и задач исследования. Принципы отбора могут включать в себя следующее:

1) Репрезентативность: тексты должны быть репрезентативными для исследуемого языка или языковой группы. Например, в Брауновском корпусе были представлены тексты на английском языке из различных жанров и источников, чтобы обеспечить широкое покрытие языковых структур.

2) Разнообразие жанров: тексты в корпусе должны представлять различные жанры (публицистика, научные тексты, художественная литература и т. д.) для того, чтобы анализировать специфические языковые особенности в каждом из них.

3) Равномерное покрытие тематик: корпус должен включать тексты на различные тематики, чтобы обеспечить достаточное покрытие языковых вариантов и использования языка в разных сферах.

4) Контролируемые параметры: при отборе текстов может учитываться также контроль некоторых параметров, например, длина текстов, частота употребления слов, распределение по времени и т.д.

Принципы отбора текстов для других корпусов могут быть аналогичными, но могут также отличаться в зависимости от специфики задач исследования. Например, для корпуса новостных текстов может быть важным обеспечить актуальность информации, а для корпуса разговорной речи – представить различные диалекты и региональные варианты языка.

# Дополнить про броуновский

3. Классификация корпусов может быть различной в зависимости от их характеристик и назначения. Вот несколько типов корпусов:

1) Исследовательский корпус: это корпус, созданный специально для проведения исследований и анализа в языковой лингвистике или других областях. Исследовательский корпус обычно содержит тексты или данные, собранные для конкретной цели исследования, и может включать в себя различные языки, жанры и стили.

2) Статический корпус: это корпус, который не обновляется или не изменяется с течением времени. Такие корпуса обычно используются для проведения исследований или анализа текстов на определенный момент времени. Статический корпус может содержать тексты, например, старые книги, статьи или письма.

3) Параллельный корпус: это корпус, содержащий параллельные тексты на двух или более языках. Такие корпуса используются для сравнения и анализа переводов, изучения лингвистических явлений и разработки машинного перевода. Параллельные тексты могут быть текстами на исходном языке и их переводами на целевой язык, а также текстами на нескольких языках, например, в рамках межъязыковых сравнительных исследований.

1. Британский национальный корпус (BNC) содержит около 100 миллионов словоупотреблений и представляет собой корпус текстов, собранных из различных источников, таких как странные разговоры, литература, газеты и т. д. Этот корпус является репрезентативным с точки зрения широкого спектра жанров и стилей текстов, что делает его ценным для лингвистических исследований.

5. 1. Конкорданс - это алфавитно-тематический указатель поисковых запросов в корпусе текстов, позволяющий найти все употребления слова или фразы в контексте.

2. Рандомизация - это метод случайного распределения элементов, который применяется в корпусной лингвистике для создания контрольных групп и устранения влияния случайных факторов на результаты исследования.

3. Коллокация - это статистически значимая связь между словами или фразами, которая часто встречается в текстах и предполагает их совместное употребление.

4. Подмассив - это часть корпуса текстов, позволяющая провести более узкую и глубокую аналитику определенной темы или явления.

5. Парсинг - это процесс анализа и разбора предложений на составляющие элементы, который используется в корпусной лингвистике для изучения синтаксической структуры текстов.

6. Лемматизация - это процесс приведения слов к своей базовой форме (лемме), что позволяет учитывать все словоформы при анализе корпуса текстов.

7. Корпус-менеджер - это программное обеспечение, предназначенное для создания, хранения, поиска и анализа текстовых корпусов, упрощающее работу исследователя в области корпусной лингвистики.

6.

"Corpus Linguistics" на сайте Linguistic Society of America ([https://www.linguisticsociety.org/resource/corpus-linguistics](https://vk.com/away.php?utf=1&to=https%3A%2F%2Fwww.linguisticsociety.org%2Fresource%2Fcorpus-linguistics))  
На данном ресурсе представлены статьи, книги и другие источники по корпусной лингвистике, которые могут помочь исследователям в изучении данной темы.  
  
"The Corpus Linguistics Website" ([http://corpus.byu.edu/](https://vk.com/away.php?utf=1&to=http%3A%2F%2Fcorpus.byu.edu%2F))  
Этот веб-сайт представляет собой базу данных корпусов текстов на различных языках. Здесь можно найти информацию о программном обеспечении для работы с корпусами, а также ссылки на исследования в области корпусной лингвистики.  
  
"Corpora List" на сайте Lancaster University ([http://www.lancaster.ac.uk/fass/projects/corpus/index.htm](https://vk.com/away.php?utf=1&to=http%3A%2F%2Fwww.lancaster.ac.uk%2Ffass%2Fprojects%2Fcorpus%2Findex.htm))  
Этот ресурс содержит список корпусов текстов, которые доступны для исследования, а также информацию о методологии и инструментах корпусной лингвистики.  
  
"Corpus Linguistics" на сайте Oxford Research Encyclopedia of Linguistics ([https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-1074](https://vk.com/away.php?utf=1&to=https%3A%2F%2Foxfordre.com%2Flinguistics%2Fview%2F10.1093%2Facrefore%2F9780199384655.001.0001%2Facrefore-9780199384655-e-1074))  
На данном ресурсе представлен обзор основных концепций и методов корпусной лингвистики, а также актуальные исследования в этой области.

**Лаб№10**

1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | НКРЯ | BNC(English corpora) | COCA |
| Количество вхождений | **39 076** | 1882580 | 49260 |

Лидерство BNC объясняется тем, что был выбран корпус новостей с объемом более 15 миллиардов слов

2.

|  |  |  |  |
| --- | --- | --- | --- |
|  | НКРЯ | BNC | COCA |
| 1 | *Меня не трогайте! Я занемог*  *Недугом русским ― чаяньем, надеждой.*  *На зов ее, как птица на манок,*  *Я отозвался с радостью поспешной.* | [Эволюция **русской** языковой картины мира и **культура** речи](https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1214&context=rlj) | mith &; Nephew countered with its own experts, including a top deputy of the Russian |
| 2 | *«Русским словом»*  *С передовыми Зайцева статьями,*  *Ни (о позор мне!* | <https://www.washingtonpost.com/world/2024/01/16/russia-air-travel-danger-sanctions/>  civilian Russian airplanes experienced at least eight serious mechanical failures | Russian Russian First Year, Part 1... In this introduction to the Russian language, culture and history, students will develop basic skills of speaking, listening, reading, . |
| 3 | *И только вносят жизнь в своем свистящем гаме*  *Лишь русские скворцы, зимующие тут.* | Even so, he wrote, what might, à la rigueur , be acceptable for aRussian Grandmaster | 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] |

3.

|  |  |  |  |
| --- | --- | --- | --- |
| Номер примера | НКРЯ | BYC(English corpora) | COCA |
| 1 | Прилагательное | Прилагательное | Прилагательное |
| 2 | Прилагательное | Прилагательное | Прилагательное |
| 3 | Прилагательное | Прилагательное | Прилагательное |

4.

|  |  |  |  |
| --- | --- | --- | --- |
| № примера | НКРЯ | COCA | BNC |
| 1 | Относится к русской нации | Относится к русской культуре | Относится к политике России |
| 2 | Слово русского языка | Самолеты изготовленные в России | Относится к русской культуре |
| 3 | Птица России | Гроссмеймтер из России | Относится к русскому языку |

5.

Взятые слова все являются прилагательными, но лексическое значение у всех разные

6.

Эти ресурсы позволяют анализировать, проводить исследования между различными языками, диалектами и другими лингвистическими аспектами. Они очень помогают в изучении языков