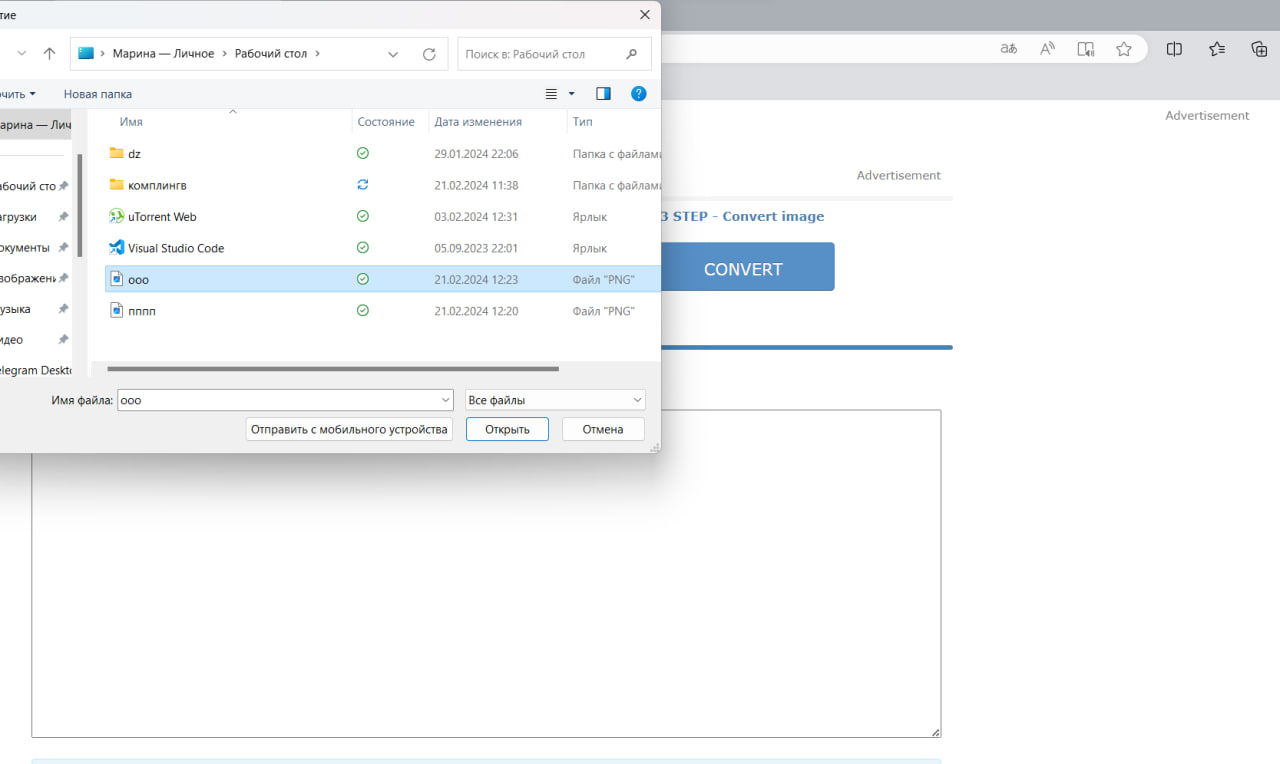
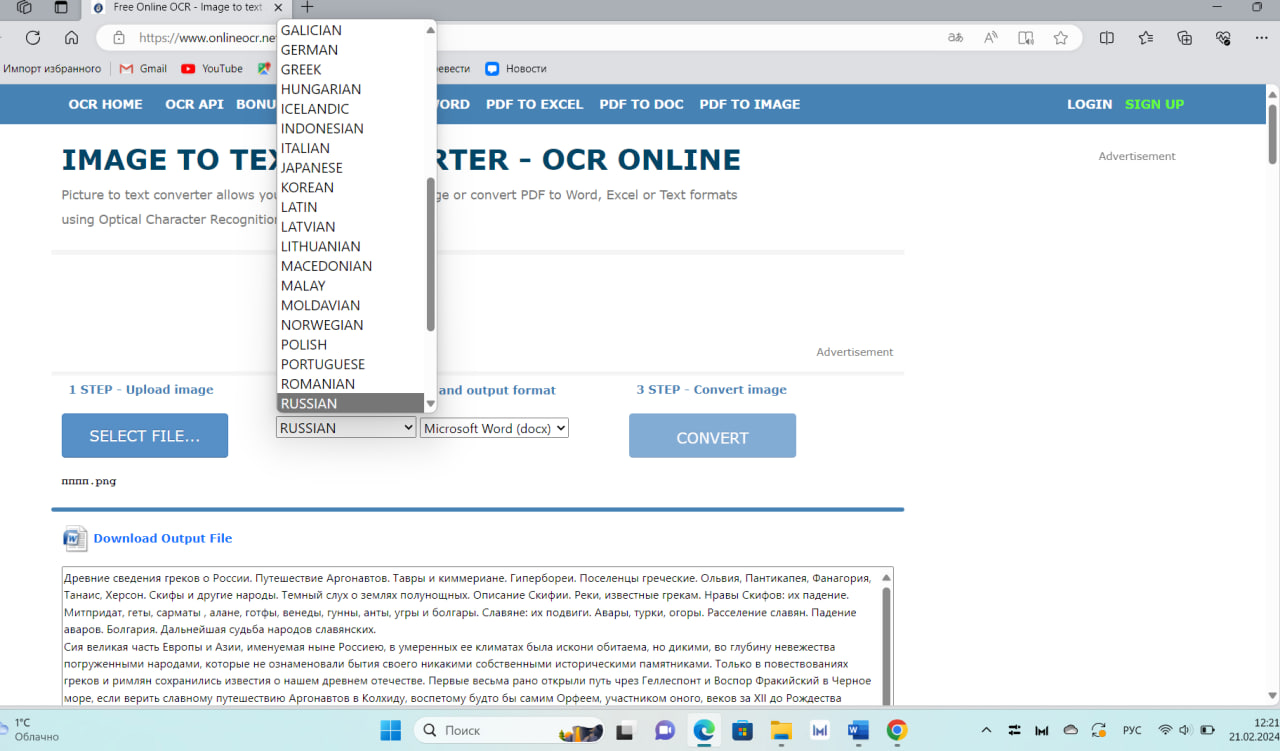
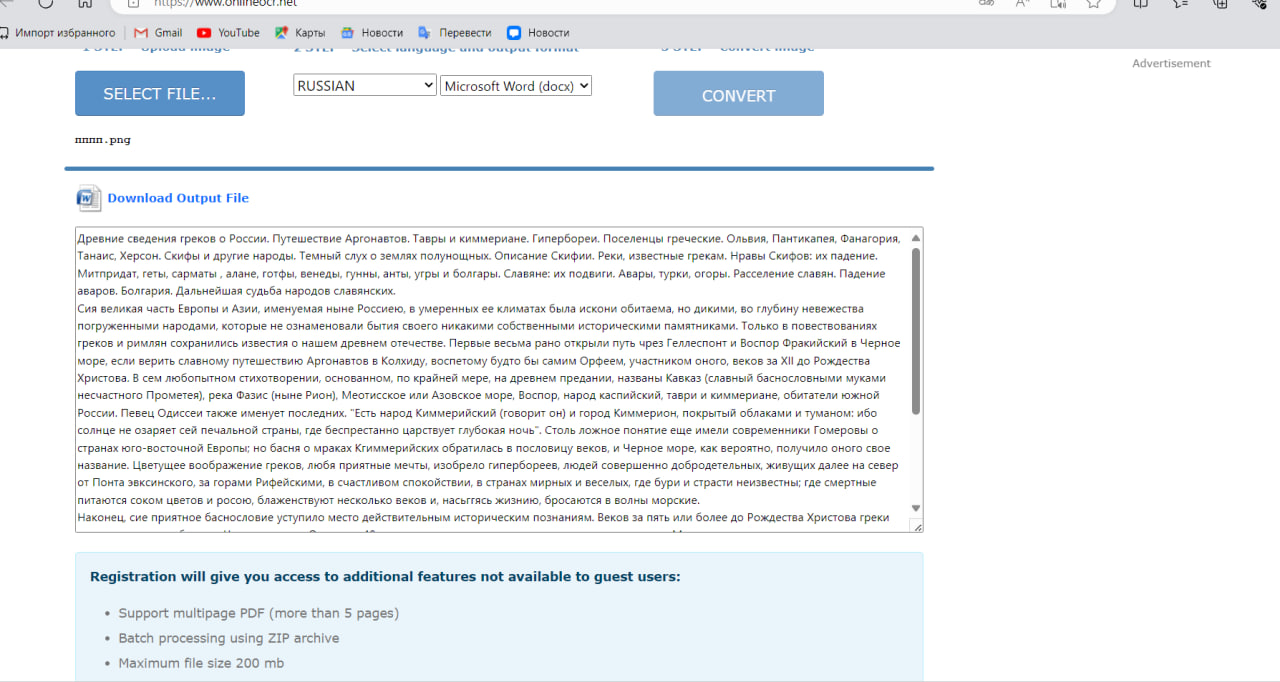
**Уровень 1**

Задание 1

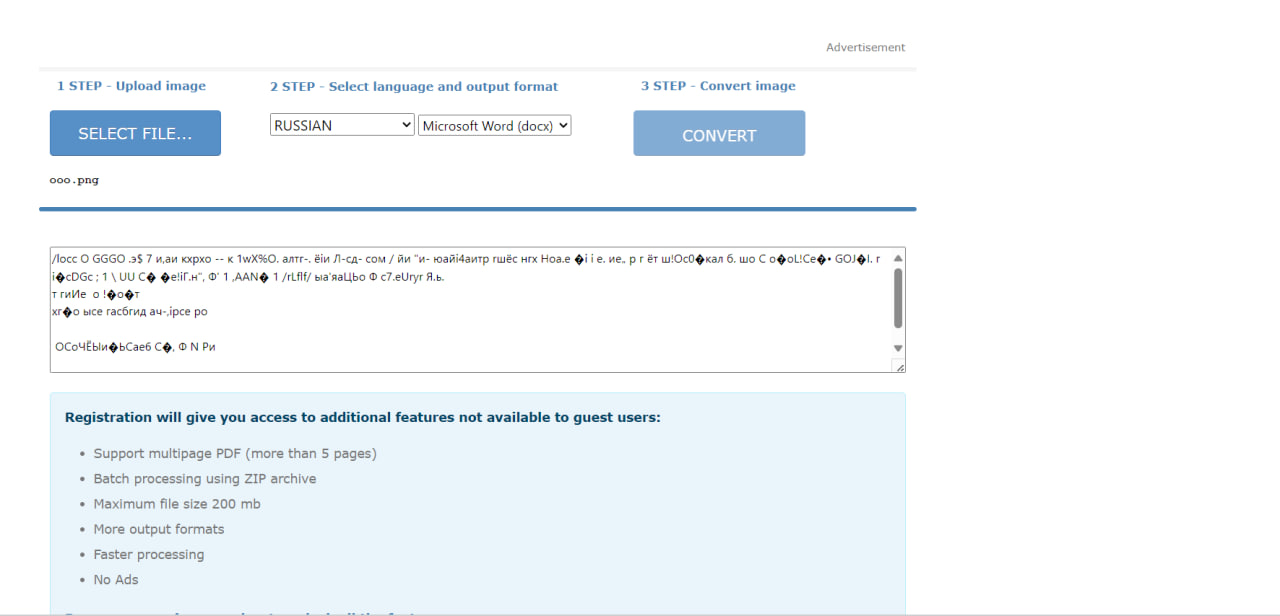
Сначала мы выбрали русский язык, вязли картинку с печатным текстом и добавили в программу.



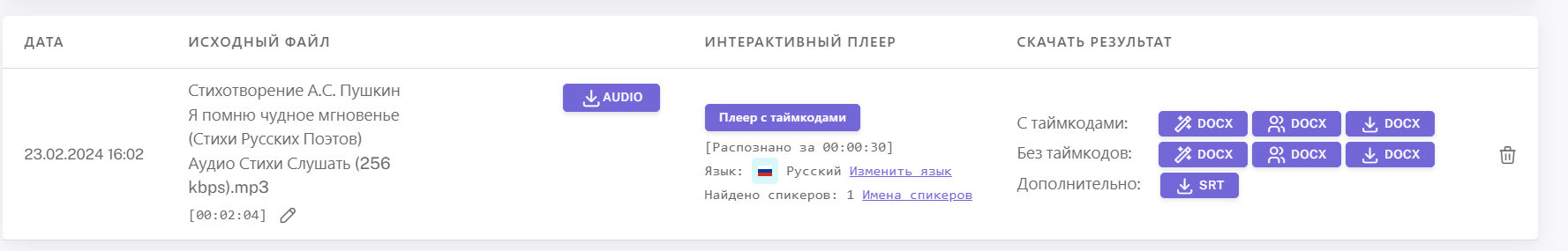
Программа без проблем перевела печатный текст с картинки.

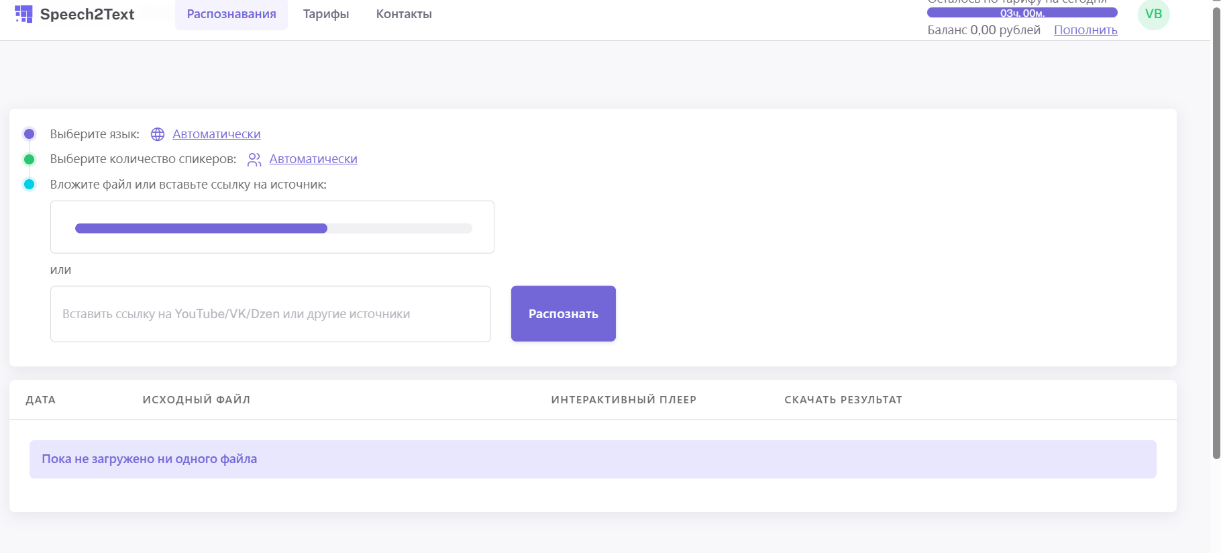


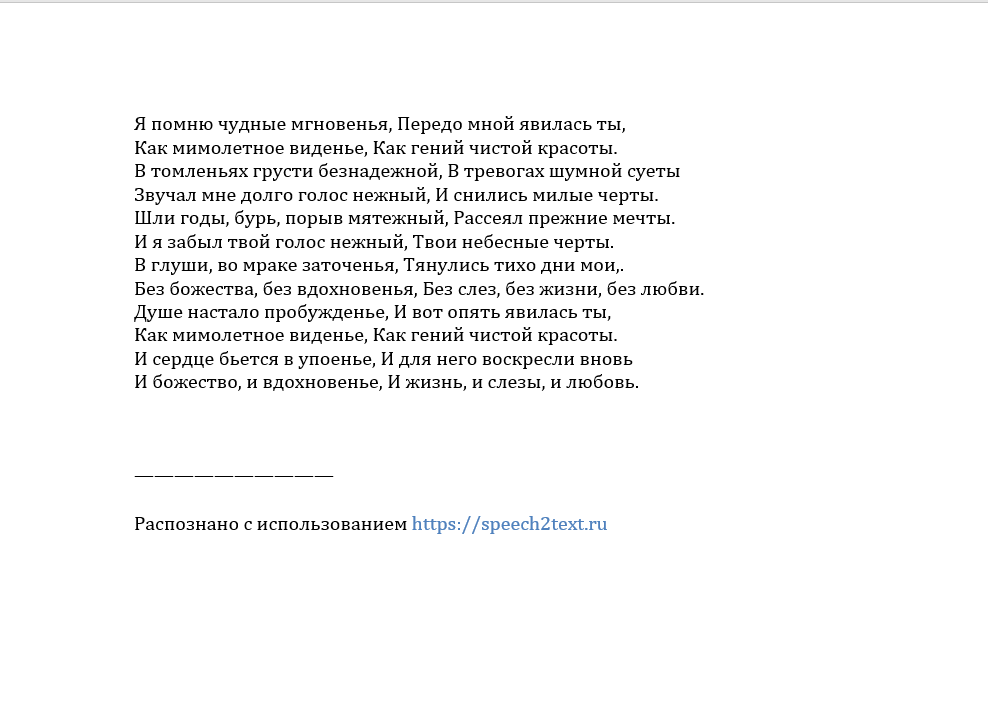
Теперь делаем то же самое, только для рукописного текста. Программа с этой задачей не справилась.



Как итог, на данный момент существует возможность переводить только печатный текст.

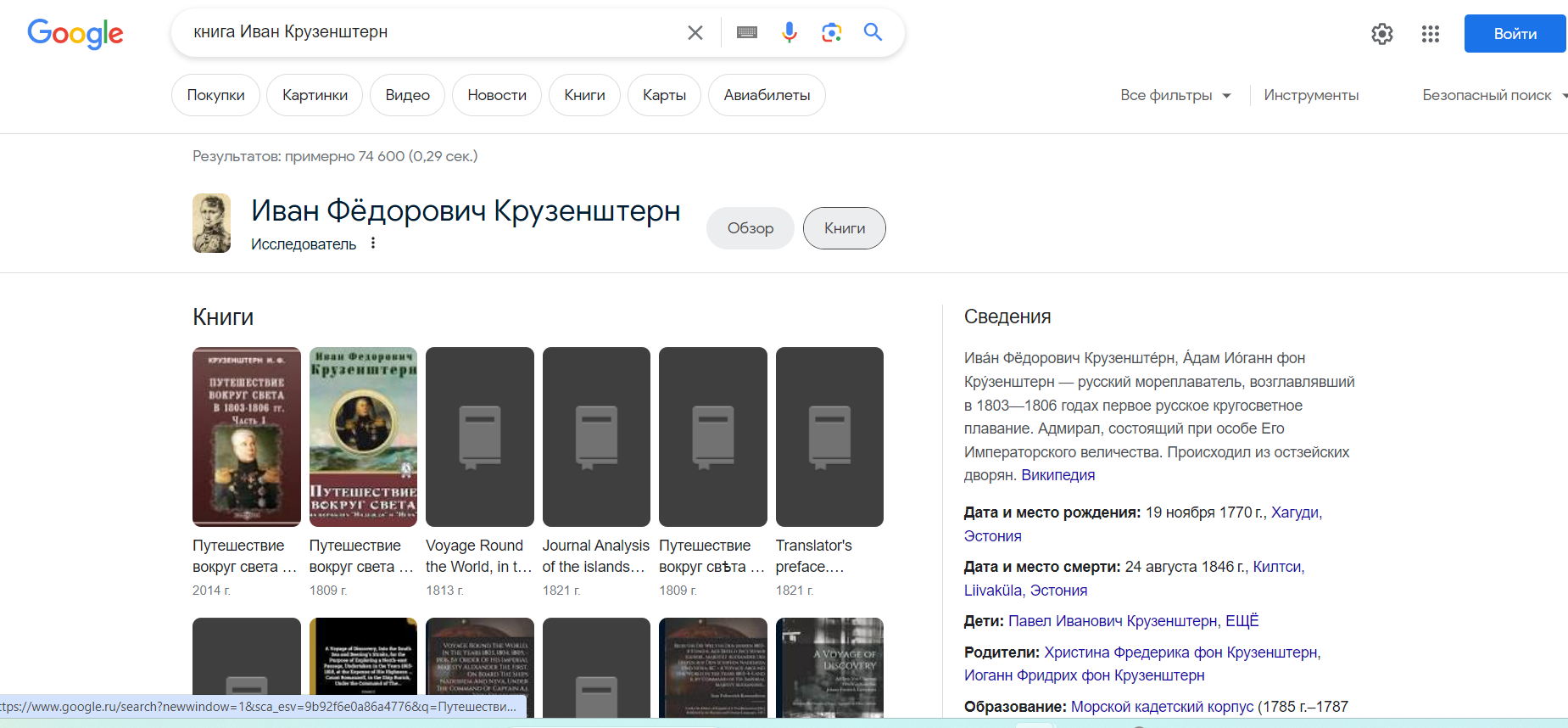
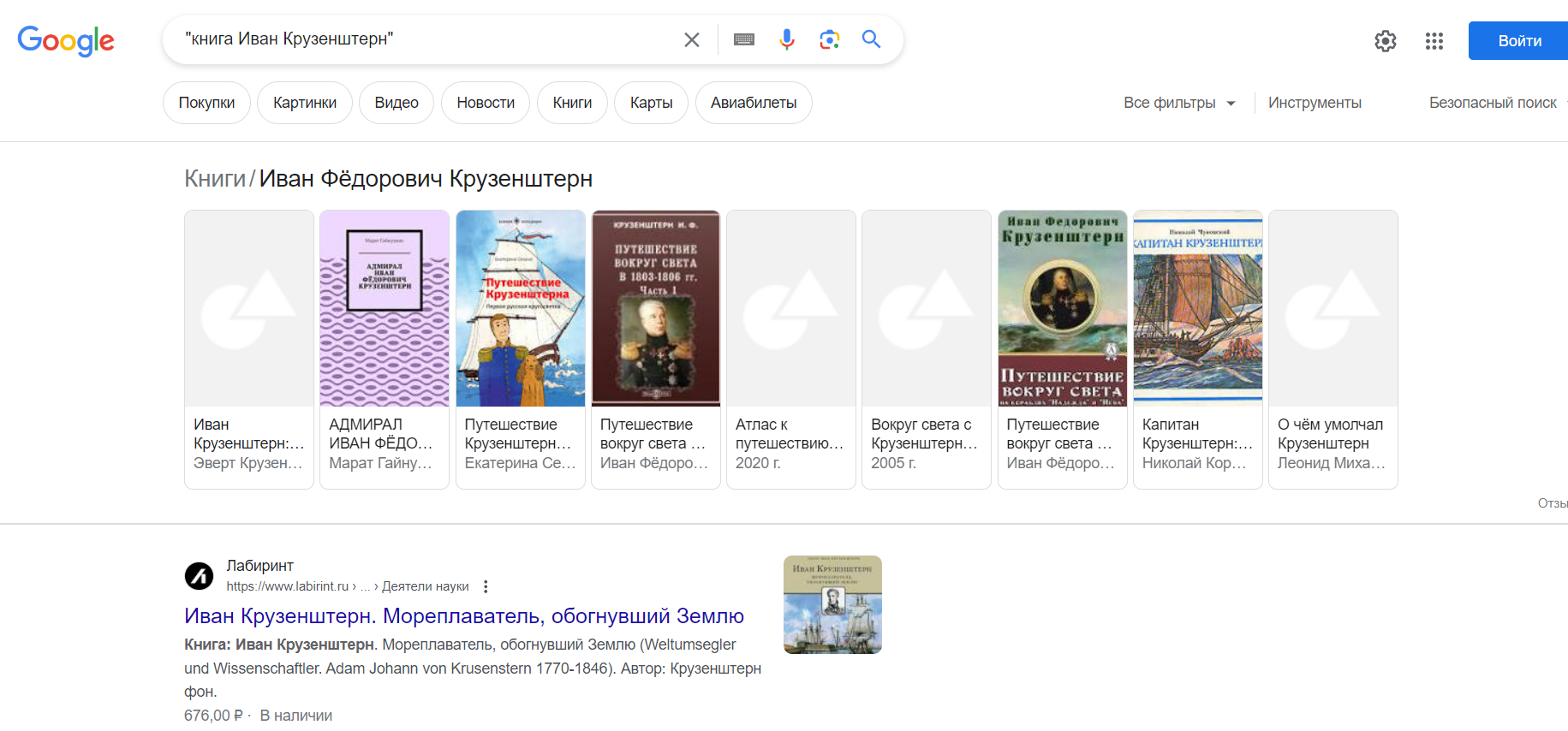
Для перевода аудиофайла в текстовый воспользуемся Speech2Text. Прикрепляем исходный файл и программа, после обработки выдает файл с текстом в формате DOCX. 

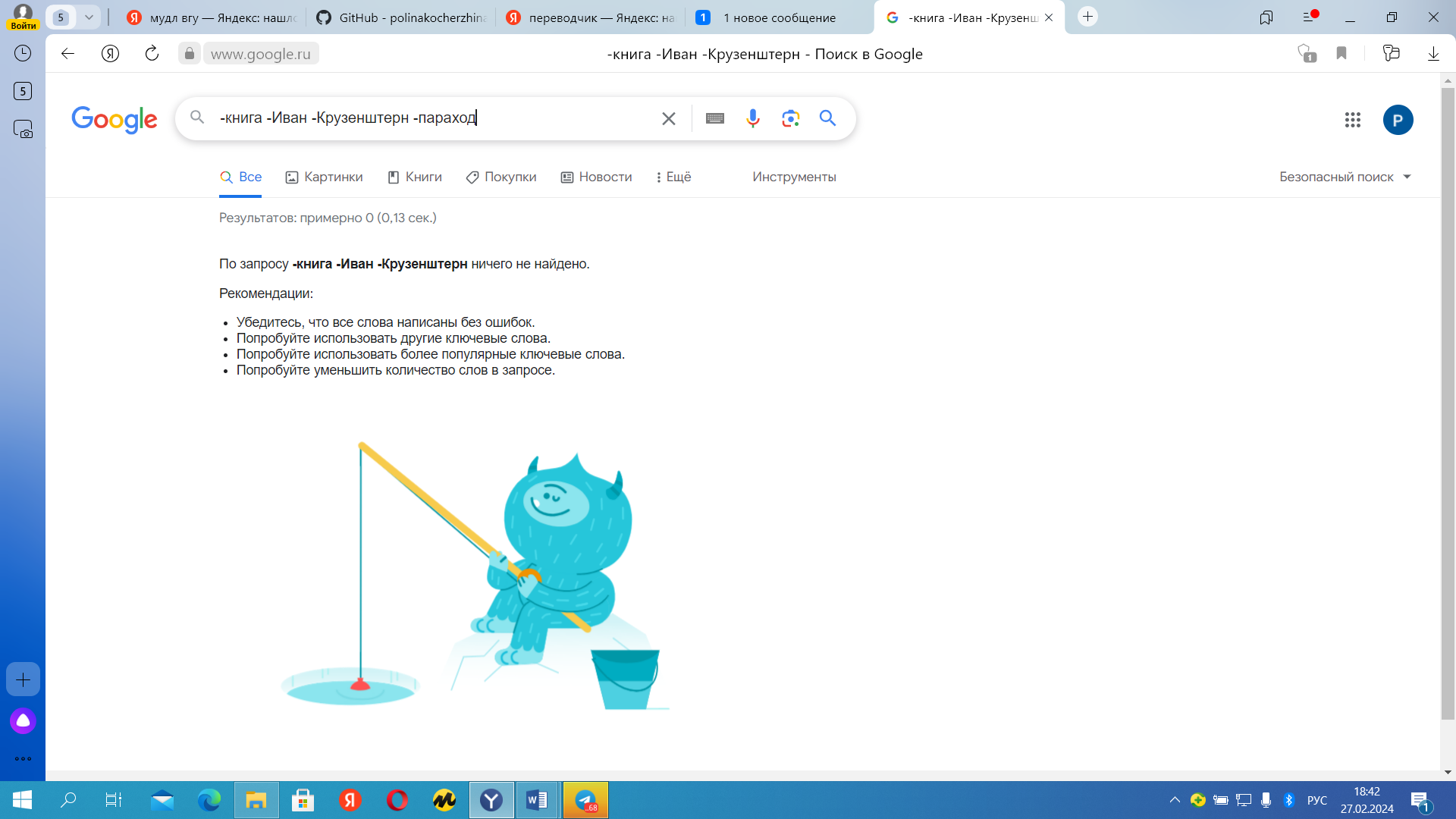


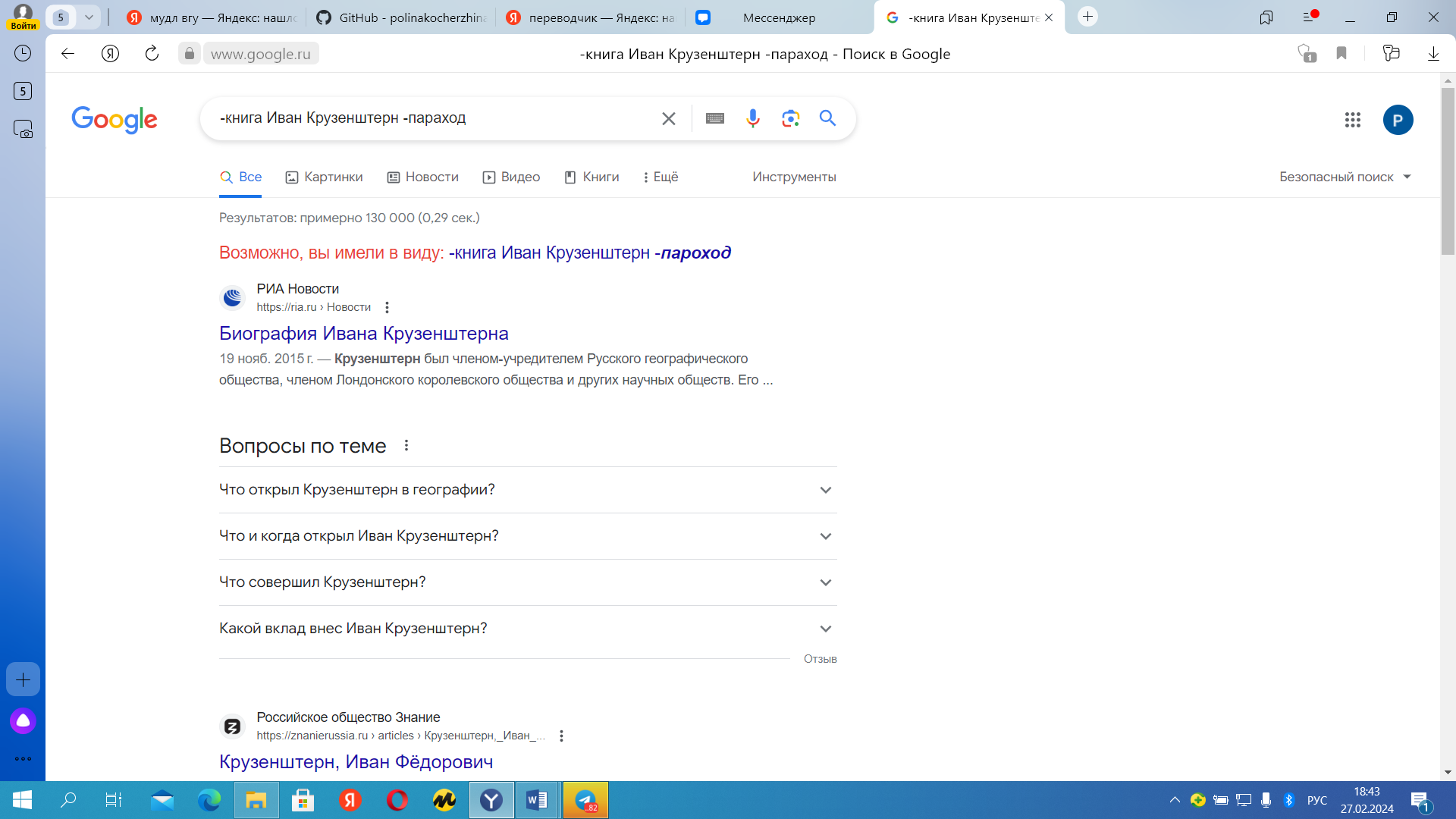
В конечном итоге получаем вот такой текстовый файл.

Задание 2

Для второго задания я использовала поисковые операторы, а также делала запросы без них. Вот что у меня получилось:

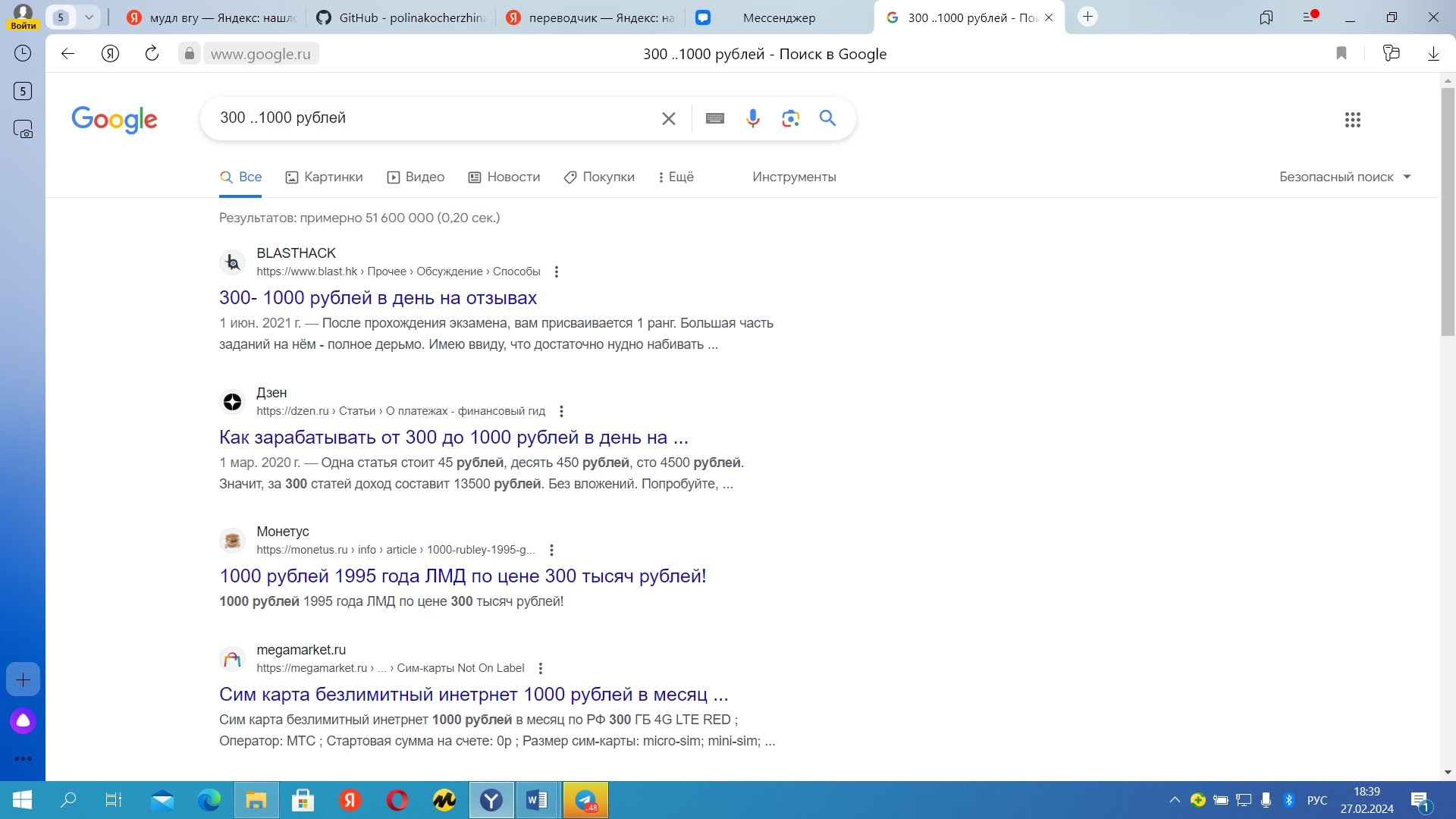
 Поисковая система вывела разные результаты. Делая запрос с поисковым оператором кавычки поисковая система гугл выдает меньше результатов, чем без кавычек, так как с кавычками она ищет точное соответствие с запросом. Если нужна точная информация, то стоит вбивать запрос с ковычками, так как кавычки дают мощные комбинации с другими операторами.



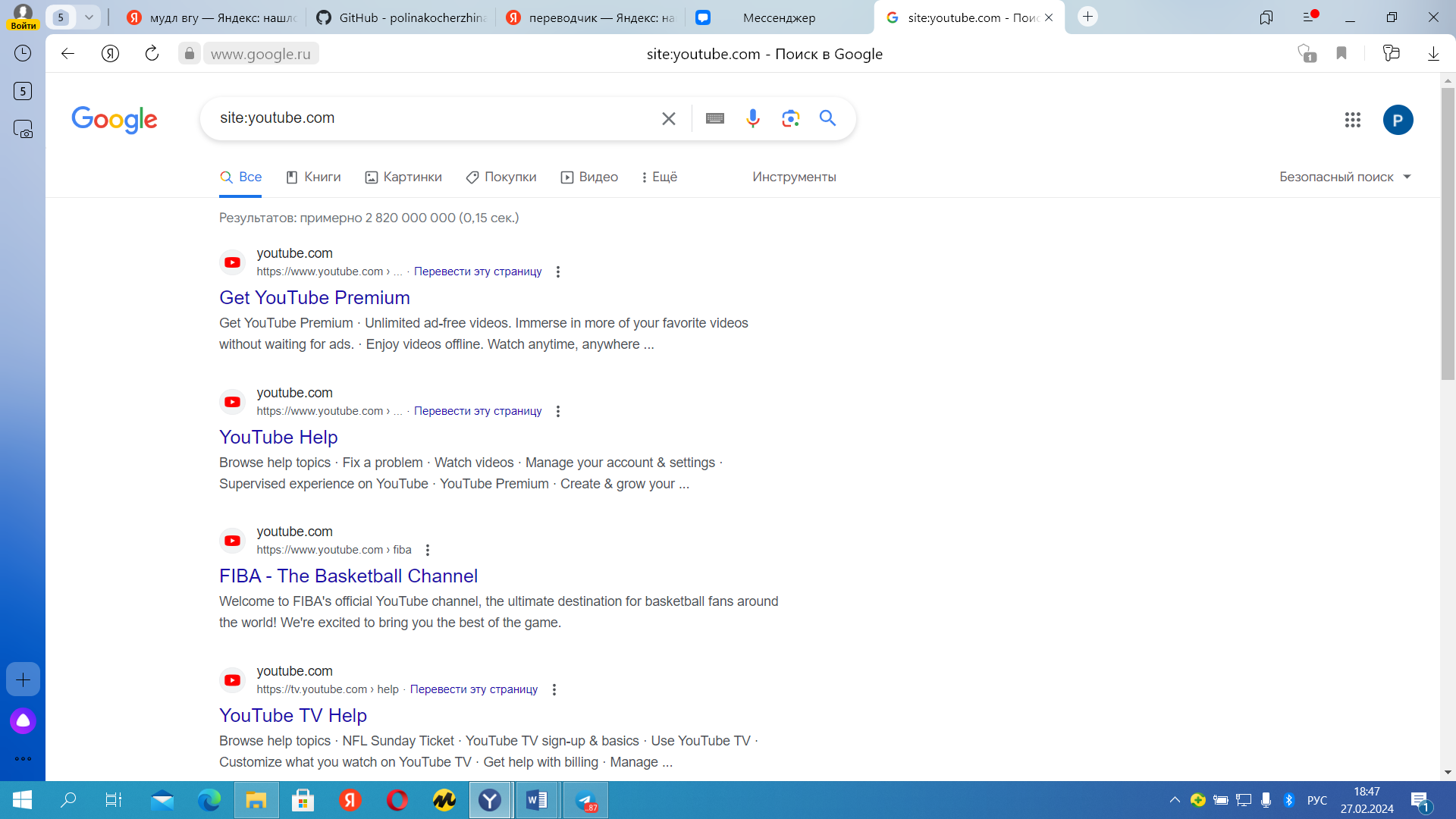
Если мы перед каждым словом ставим - , то поисковая система ничего не выдвёт, но если же перед самой фразой ставим - , то запрос выдаётся хорошо.



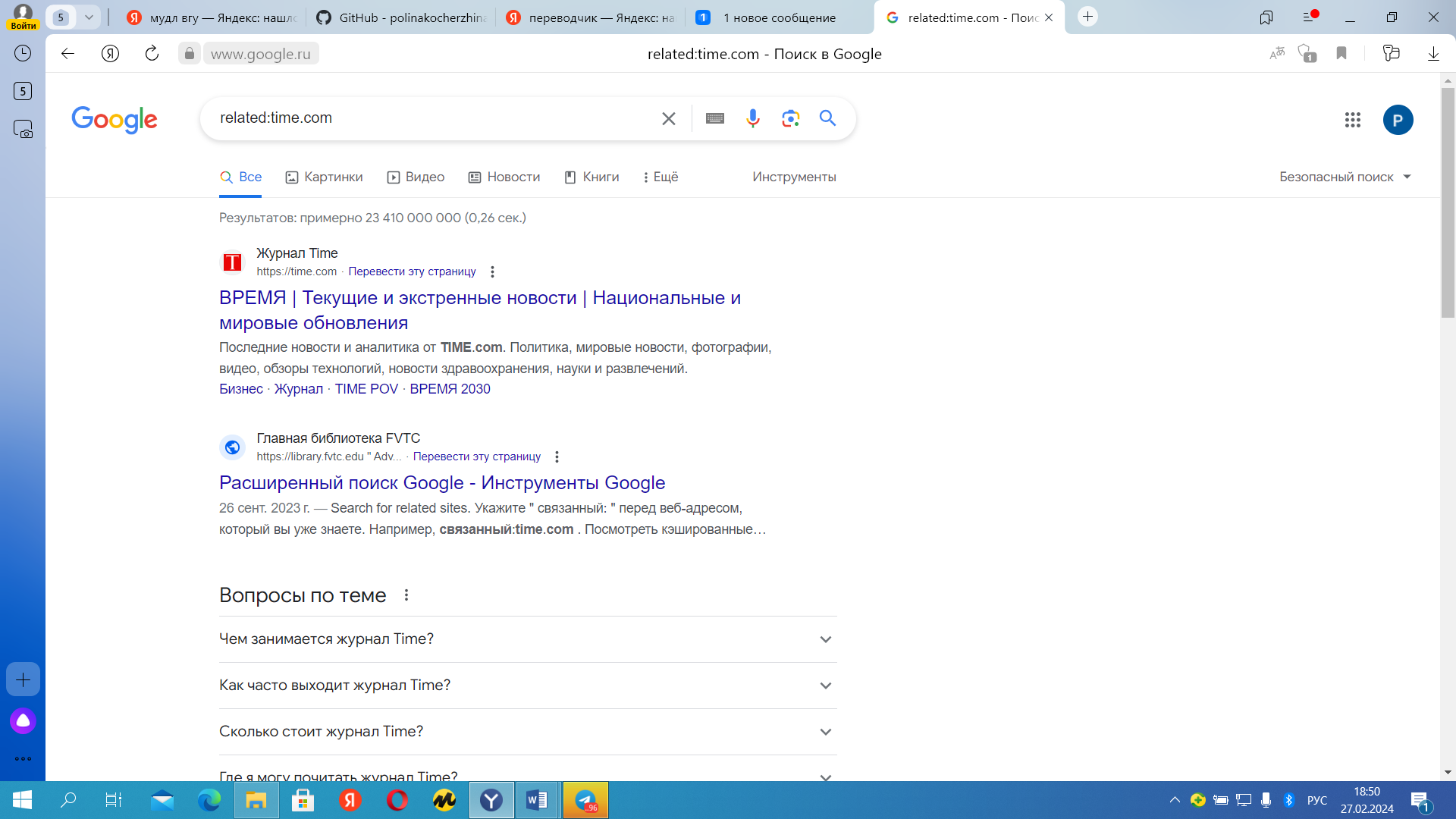
При вводе в Яндекс между числами точек, они автоматически заполняются нулями и при указании единицы измерения – рубль нам переводится рубли в доллары, если же указать другую единицу измерения, то и перевод будет в другую валюту. Но точки между цифрами в любом случае будут заменены нулями.



При вводе такого же запроса в Гугле мы можем наблюдать что выдаётся совсем другой ответ. Тут нам представляется заработок.



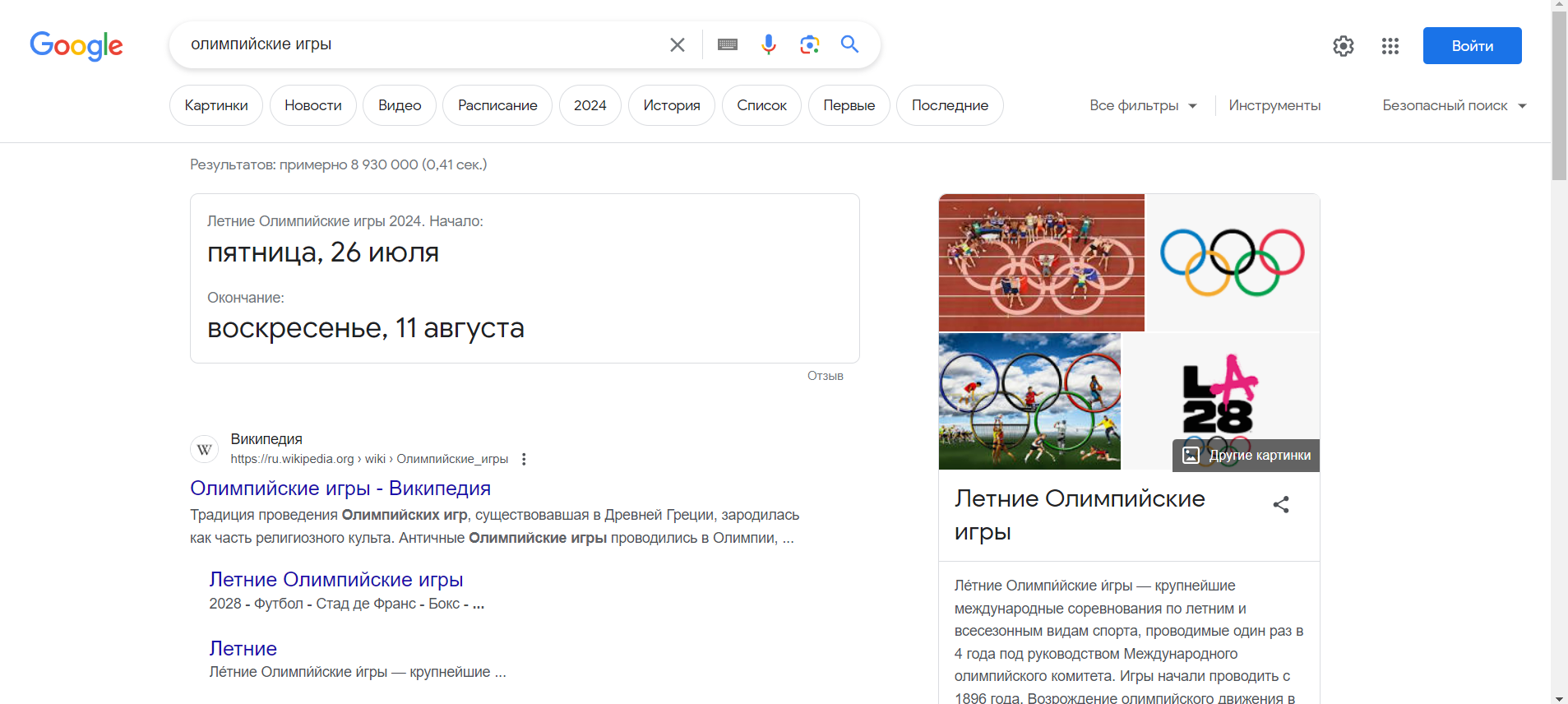
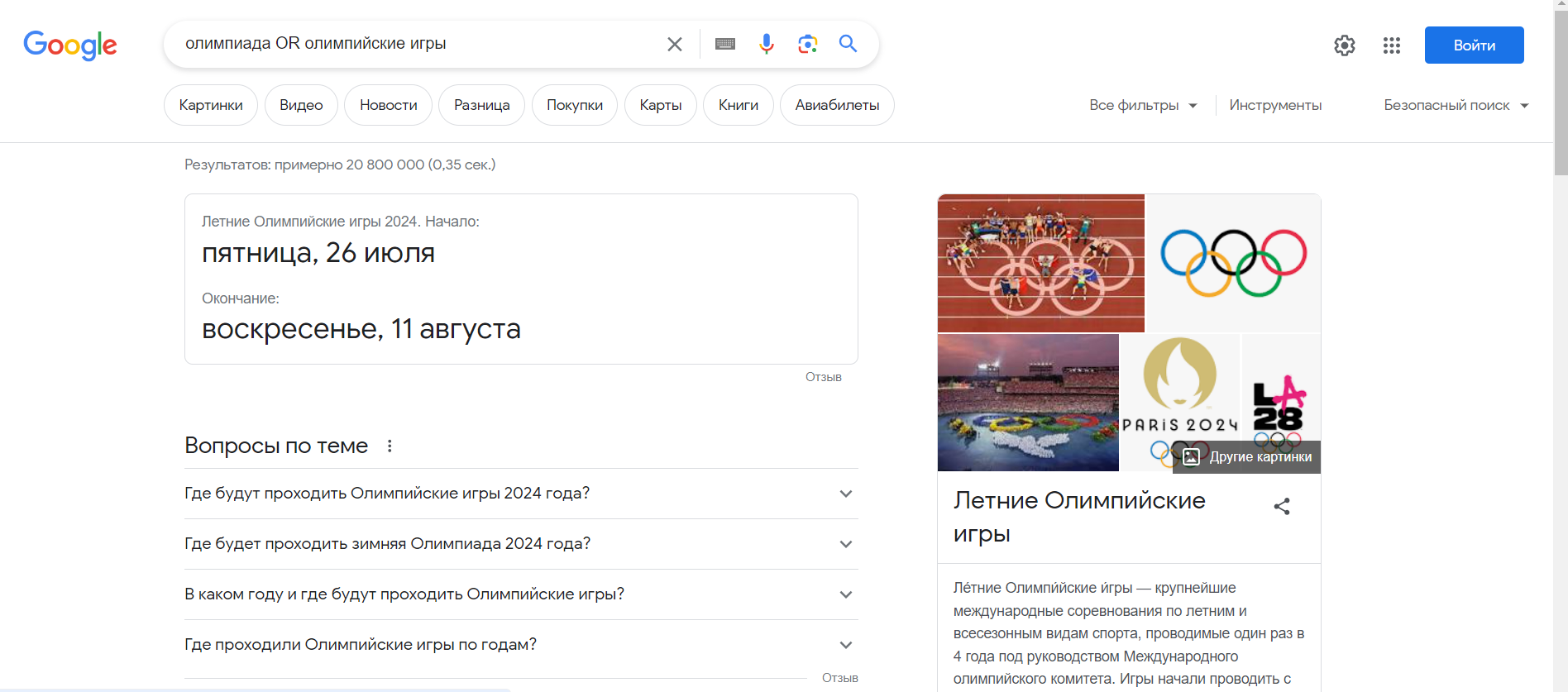
При вводе определённой ссылки поисковая система выдаёт нам конкретный запрос.



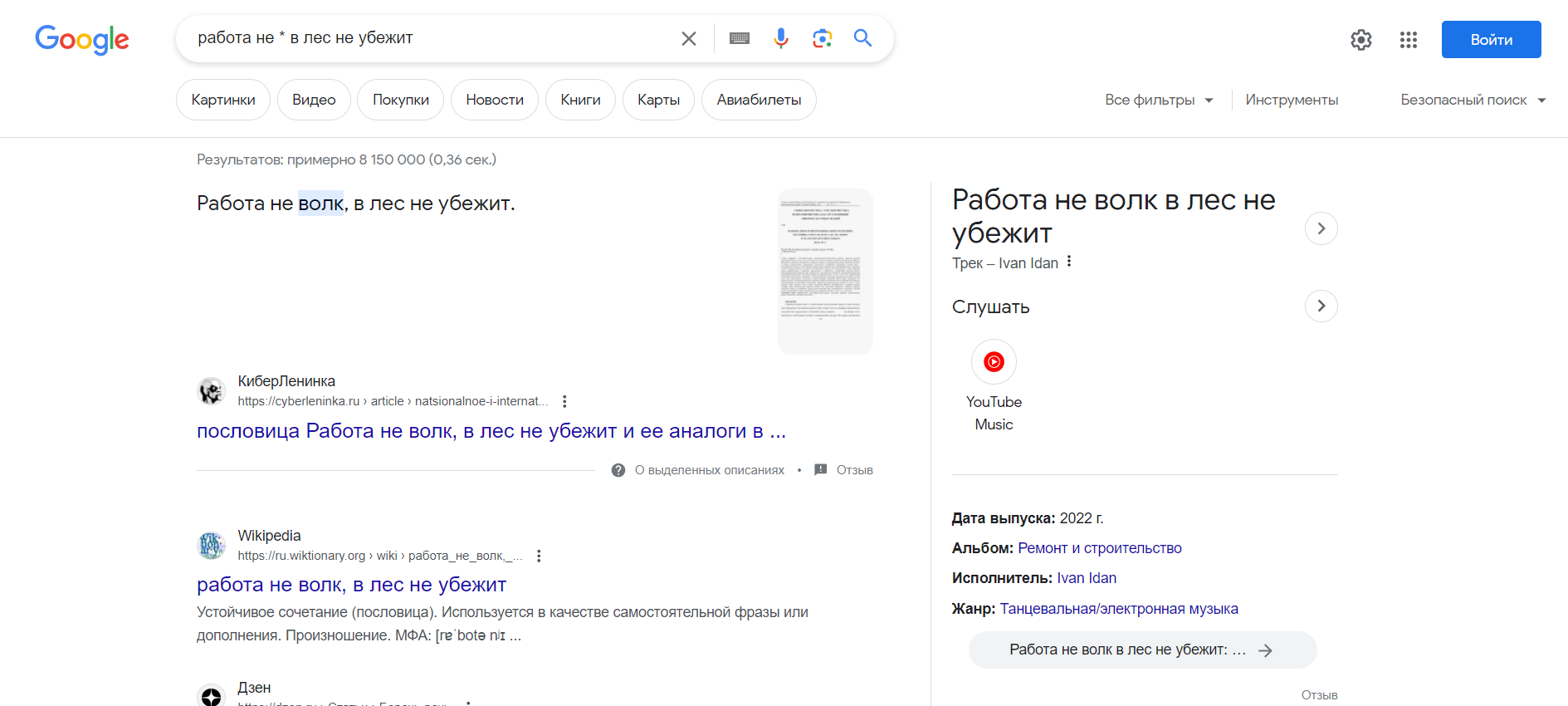
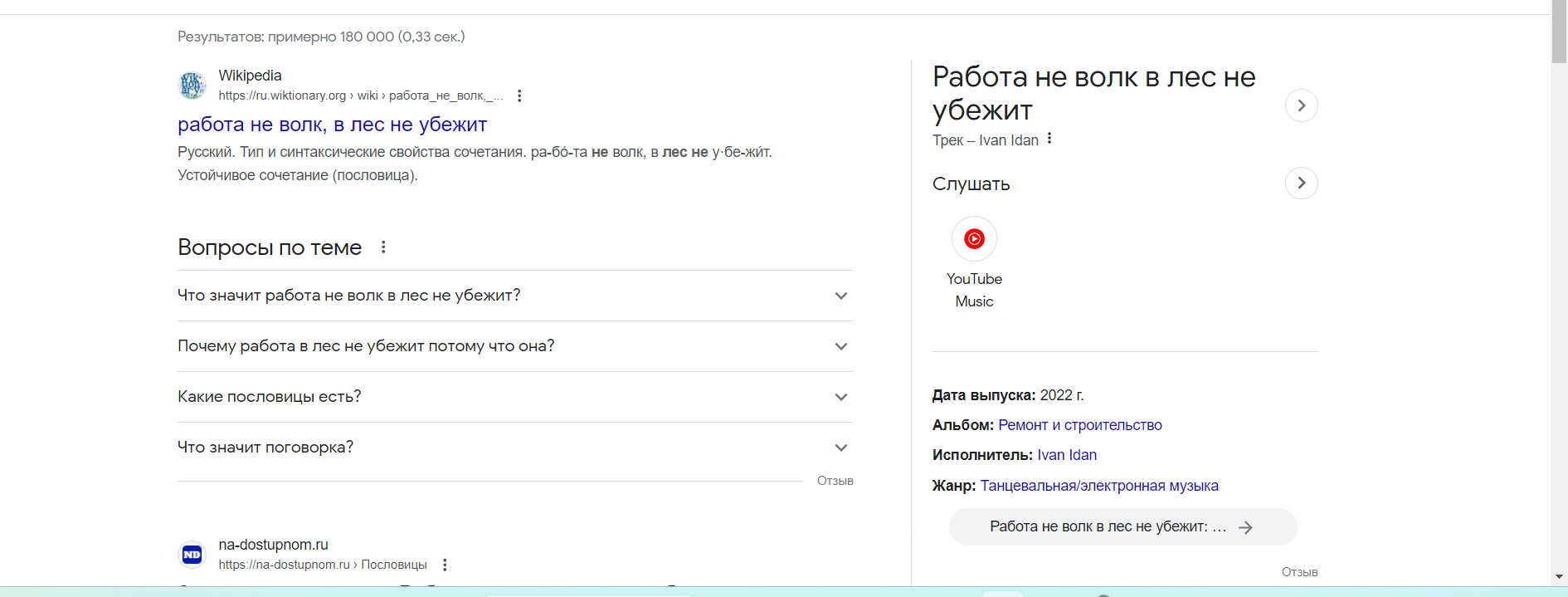
При вводне неизвестных ссылок мы получаем разные сайты, которые приближённые к запросу.



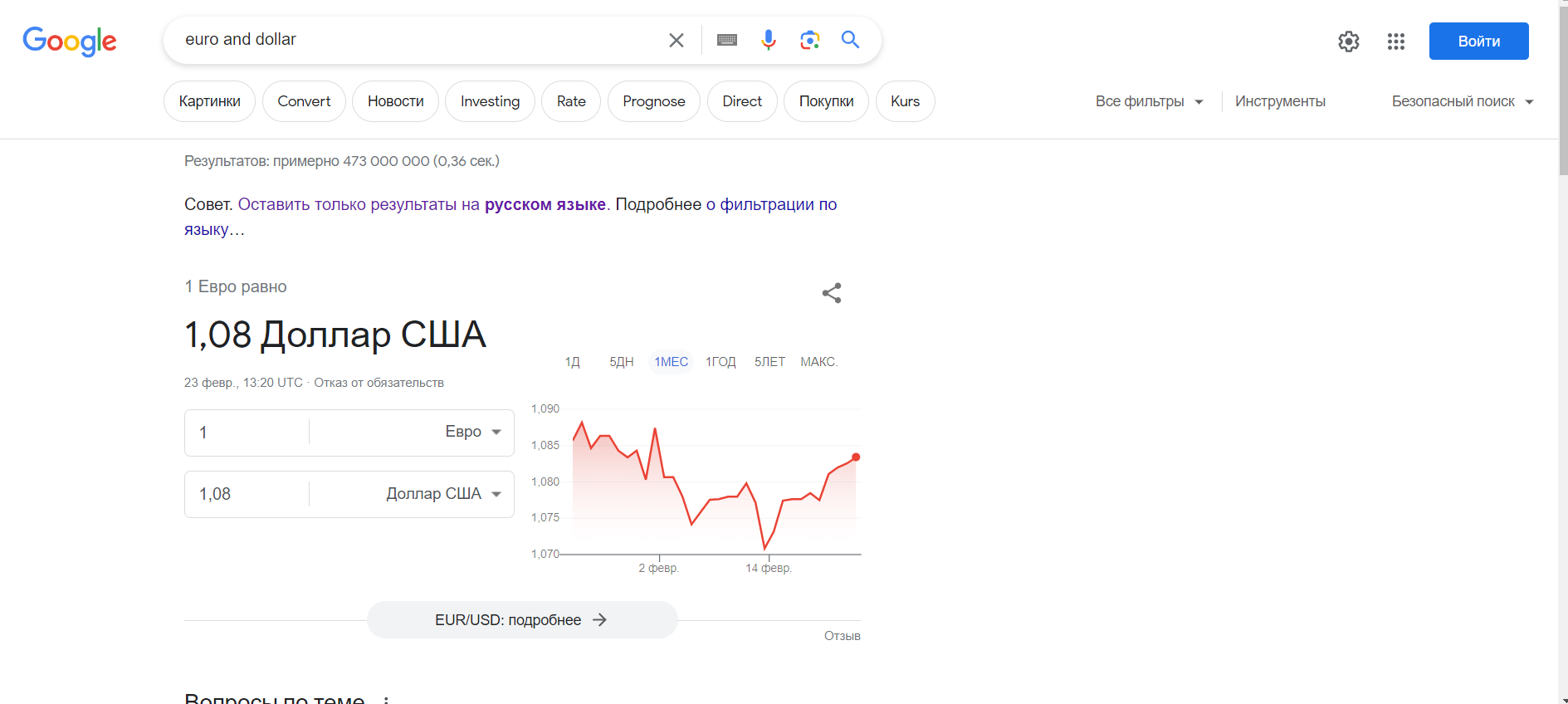
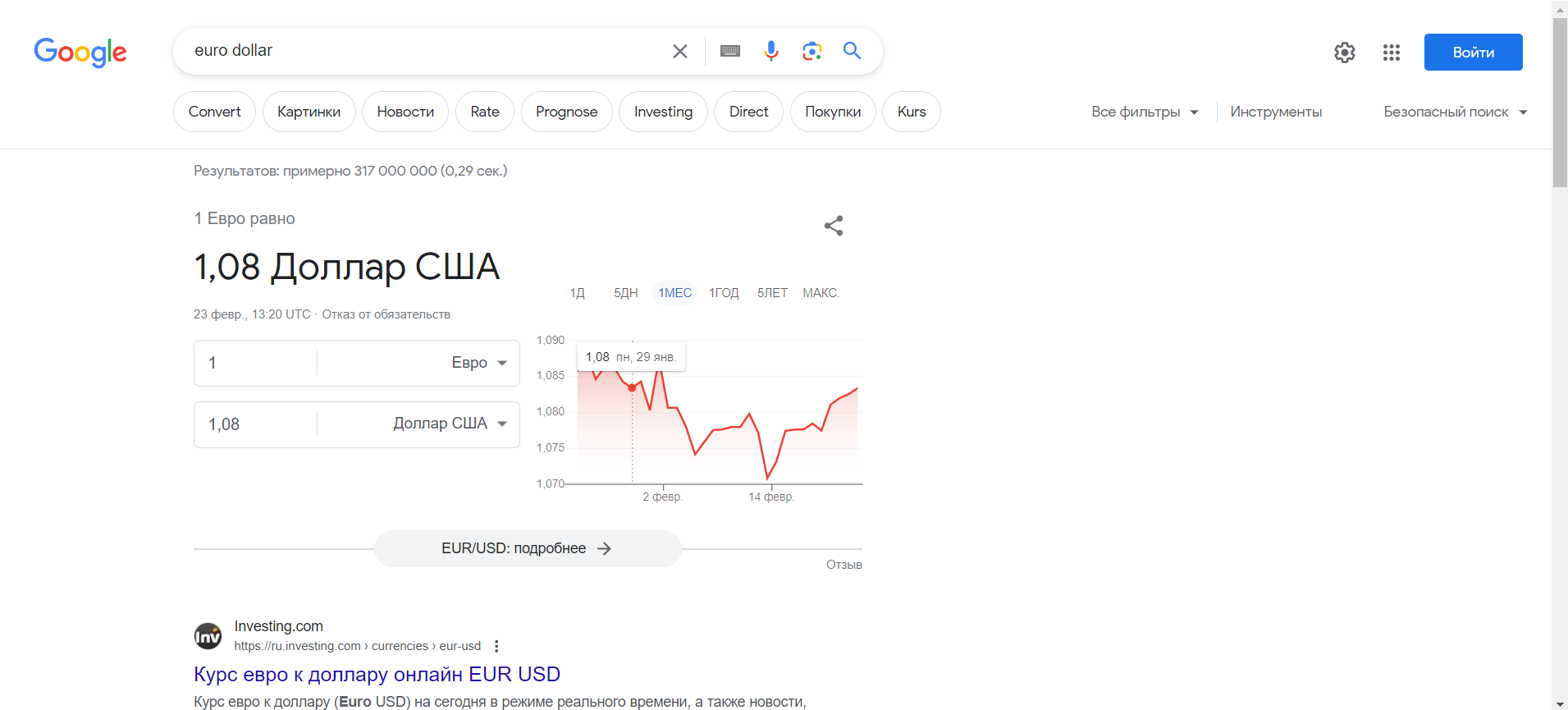
Формат файлов позваляет найти сайты с этим форматом.



Поисковй оператор OR позволяет найти объединение запросов. Следовательно, с ним результатов будет больше, чем без него.

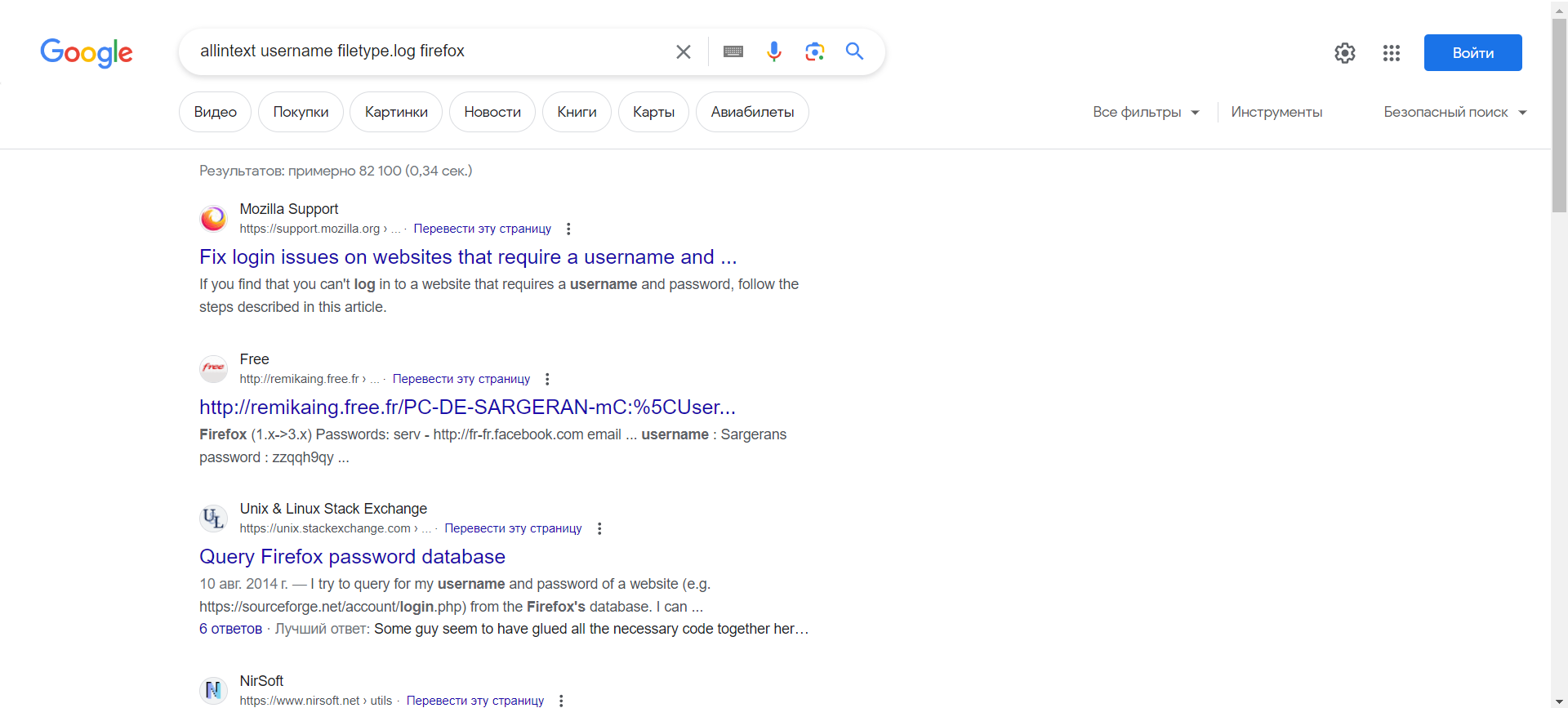
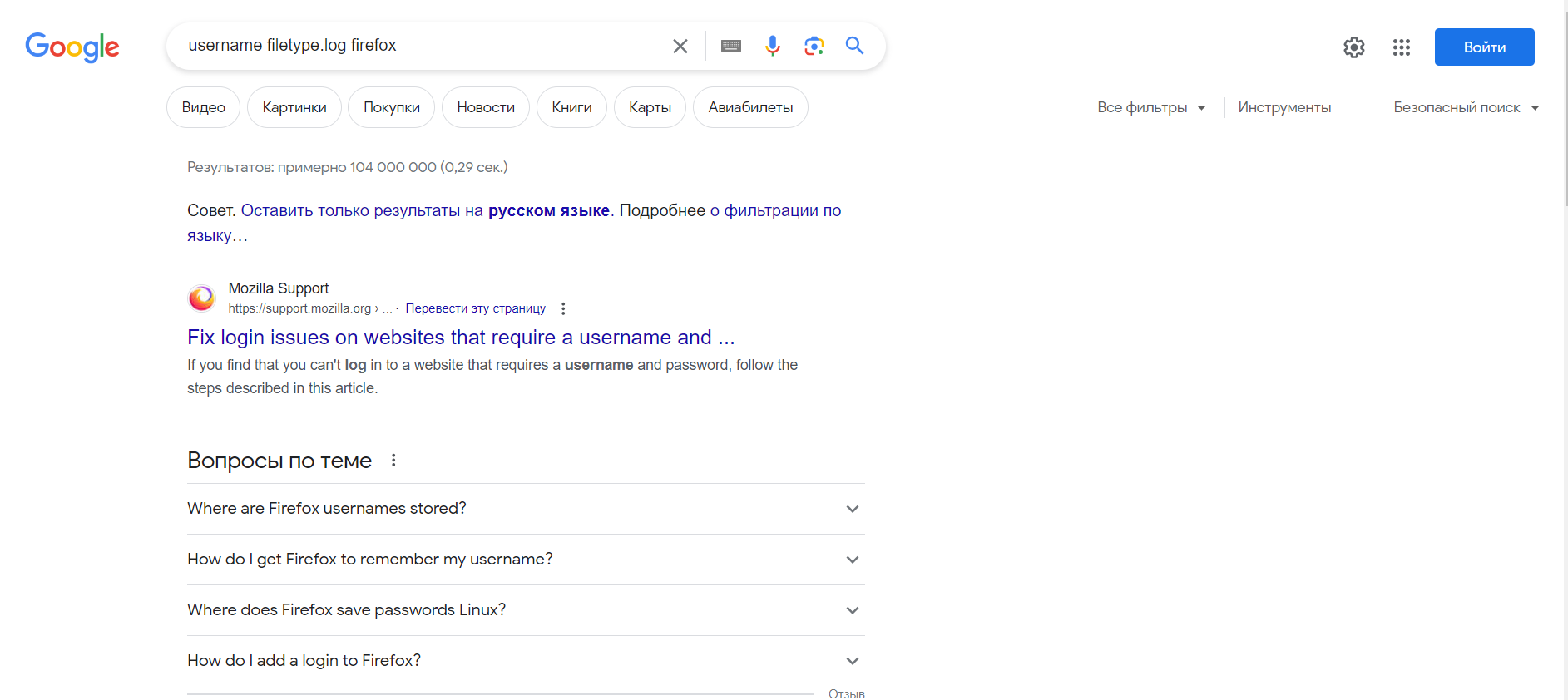
 

Далее я использовала \*. С ней поиск более расширенный и точный. Позволяет восстановить пропущеное слово. Поиск и без поискового оператора справляется с этим.

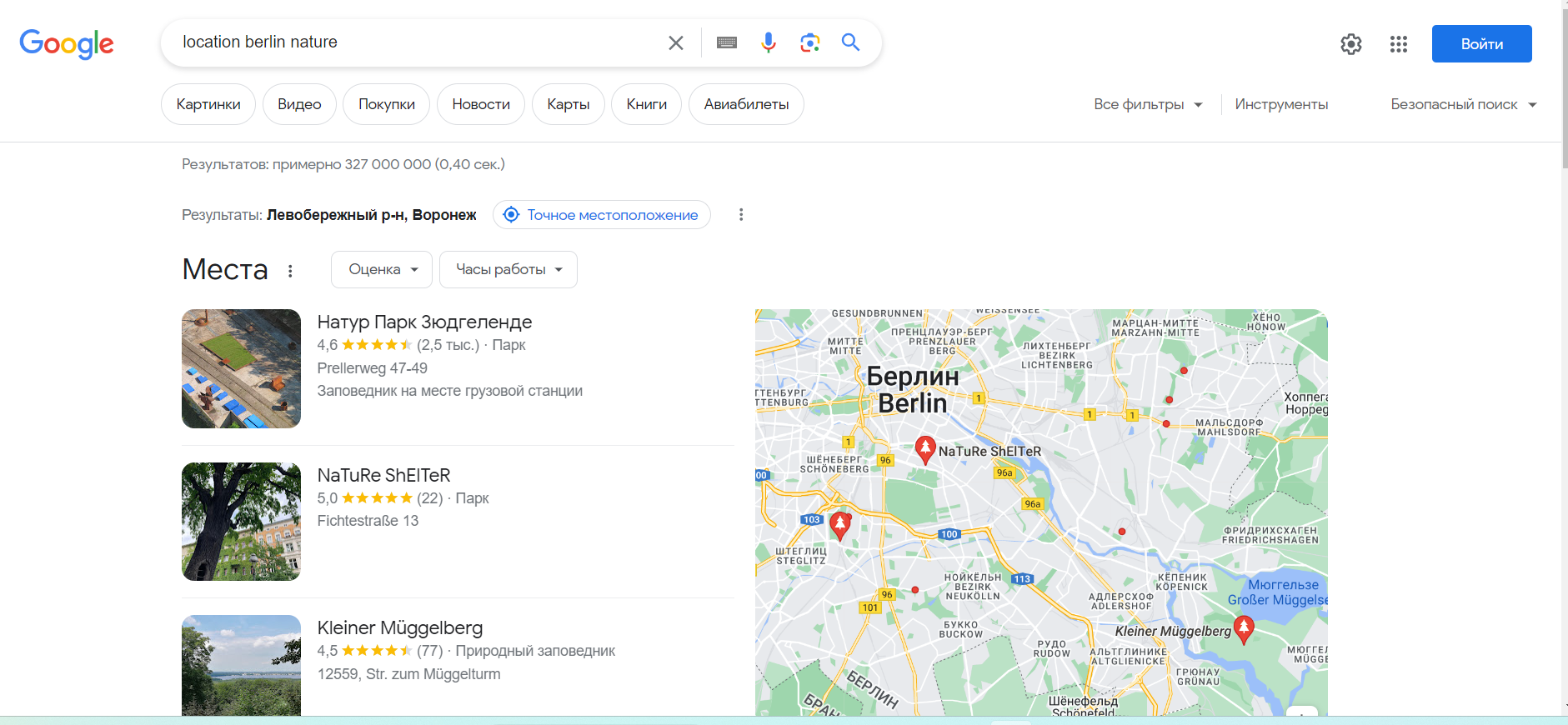
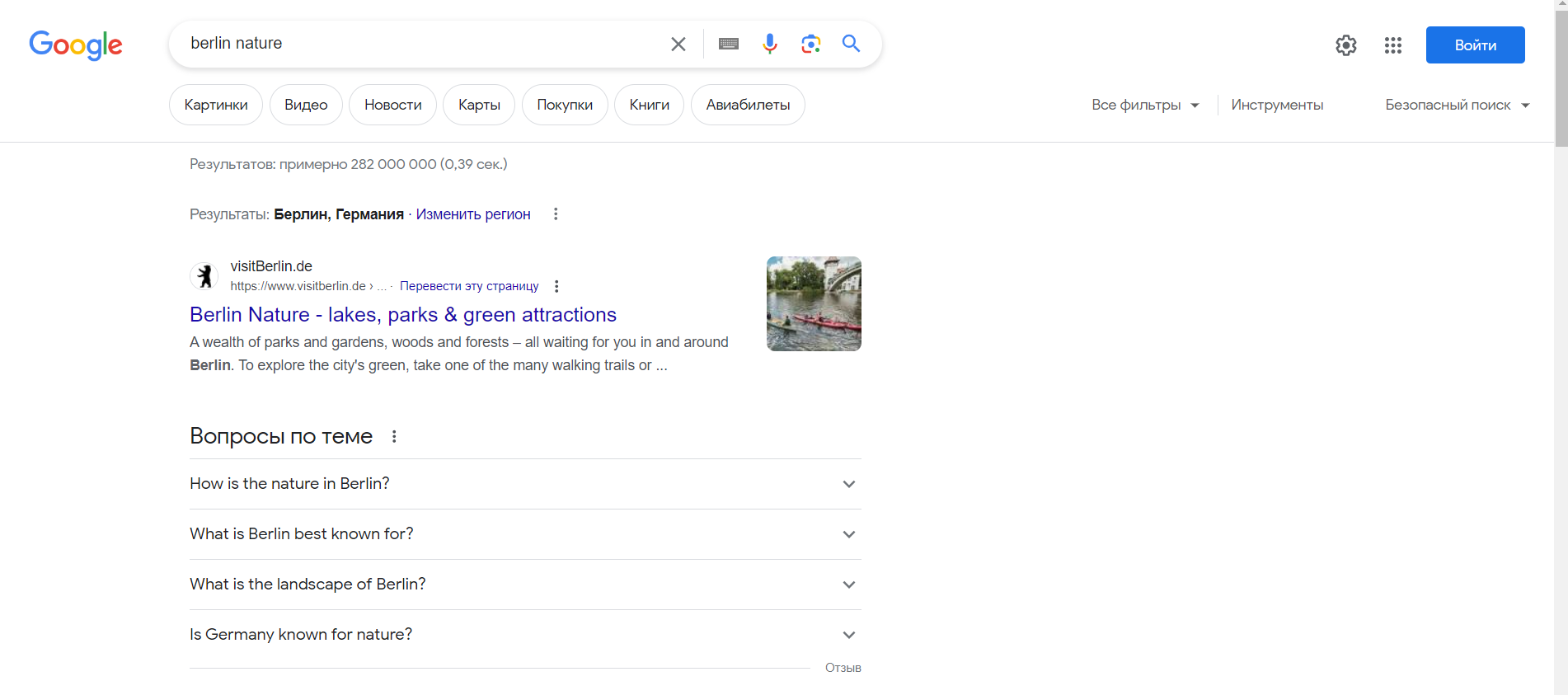
 

Поисковый оператор AND отображает результаты всех ключей (их может быть 2 и больше), которые объединены этим оператором. С ним поиск также более расширенный, чем без него.

Поисковый оператор allintext находит страницы, содержащие все указанные слова в тексте. Может быть полезен, когда нужно сделать запрос более точным.

Location ищет новости в Google News с учетом указанной локации.

|  |  |  |
| --- | --- | --- |
| Название | Особенности | Скриншоты |
| [**Google Schol****ar**](https://scholar.google.com/) | Поисковая система по полным текстам научных публикаций всех форматов и дисциплин  Позволяет экспортировать цитаты и библиографические ссылки в различных форматах  Поддерживает язык поисковых запросов Google |  |
| **Cyberleninka.ru** | научная электронная библиотека, построенная на парадигме [открытой науки](https://cyberleninka.ru/article/n/vozmozhnye-puti-razvitiya-otkrytoy-nauki-v-rossii) (Open Science)  Основными задачами является популяризация науки и научной деятельности, общественный контроль качества научных публикаций, развитие междисциплинарных исследований, современного института научной рецензии, повышение цитируемости российской науки и построение [инфраструктуры знаний](https://cyberleninka.ru/article/n/infrastruktura-otkrytoy-nauki). |  |
| [**Academia.edu**](https://www.academia.edu/) | Позволяет искать статьи, выложенные самими авторами в открытый доступ.  используется для того, чтобы делиться с другими своими статьями, отслеживать их цитируемость и следить за новостями исследований и разработок. |  |
| [**Scopus**](https://www.scopus.com/) | Библиографическая и реферативная база данных и инструмент для отслеживания цитируемости статей  Данная база данных должна стать наиболее полным и исчерпывающим ресурсом для поиска научной литературы (согласно стратегии компании)  Поиск по направлениям, фильтрация по квартилям |  |
| [**Web of Science**](https://webofknowledge.com/) | Базы данных публикаций в научных журналах и патентов  Статьи учитываются преимущественно на английском языке  Акцент на следующие направления: медицина, биология, психология, экономика, физика, астрономия; в меньшей степени математика, юриспруденция, политология и компьютерные науки |  |
| [**eLIBRARY**](https://elibrary.ru/) | Российская научная электронная библиотека  Насчитывает более 30,7 млн статей  Поиск по автору, направлению, ключевым словам  Сортировка по годам, по цитируемости  Некоторые тексты есть в открытом доступе |  |

Существует большое количество баз источников, позволяющих найти абсолютно любую информацию в интернете. Так же они позволяют найти информацию из любой точки мира, так как работают на разных языках.

**Уровень 2**

Задание 2

1. **Основные возможности современных систем распознавания (OCR):**

**- Получение текстового документа** на основе отсканированного изображения.

**- Получение базовых атрибутов документа**. Это позволяет облегчить процесс занесения данных в информационные системы при регистрации документов.

**- Классификация документов**. Система распознавания определяет тип документа.

**- Распознавание неструктурированных документов**. Системы способны полностью систематизировать и атрибутировать документы.

**- Определение комплекта документов**, объединённых одной логической связью.

**- Работа со справочниками**. Системы распознавания могут исправлять ошибки и неточности в распознанных данных.

**- Единый информационный ландшафт**. Распознанный документ формирует новую электронную карточку или привязывается к действующей карточке информационной системы.

2. По данным источника, перспективы развития OCR-программ включают:

**- Использование глубокого обучения**. Современные OCR-инструменты используют нейронные сети для распознавания символов и понимания контекста.

**- Расширение возможностей многоязыковых OCR-систем**. OCR может обрабатывать различные языки, устраняя языковые барьеры.

**- Использование дополненной реальности (AR) и OCR**. Сочетание AR с OCR может привести к мгновенному переводу и пониманию текста.

**- Обработка данных в реальном времени и мгновенный захват данных**. OCR может использоваться для обработки данных на лету, особенно в сферах, таких как банковское дело и здравоохранение.

**- Улучшение безопасности и конфиденциальности данных**. Будущие OCR-инструменты могут использовать передовые методы шифрования и децентрализованный подход для обработки данных на устройстве пользователя.

3. Интеллектуальное распознавание символов - это усовершенствованное оптическое распознавание символов или, более конкретно, система распознавания рукописного ввода, которая позволяет компьютеру изучать шрифты и различные стили почерка во время обработки для повышения точности и уровня распознавания.

4. Особенностях системы оптического распознавания текста от ABBYY:

- Быстрота и точность распознавания.

- Полное сохранение исходной структуры и форматирования документа.

- Поддержка более 190 языков.

- Распознавание символов, набранных любым шрифтом.

- Возможность сохранения текста почти во всех редактируемых форматах (DOC, TXT, RTF, XLS, HTML, PDF), автоматической передачи документа в другие приложения.

- Автоматизация однотипных операций, что позволяет распознавать и обрабатывать документы еще быстрее.

Задание 3

Лабораторная 1

|  |  |
| --- | --- |
| [Задание](https://edu.vsu.ru/mod/url/view.php?id=919316) | Ответ |
| 1. Сколько раз в тексте встречается слово child (в разных формах)? | 22 |
| 2. Сколько раз в тексте встречается слово child именно в этой форме? | 10 |
| 3. Приведите один из контекстов использования в тексте слова beautiful | … and THEN-she found herself at last in the beautiful garden.. |
| 4. В какой орфографии (британской или американской) представлен текст? | Ответ: британской  Проверочное слово: pardon |

Лабораторная 2

|  |  |  |
| --- | --- | --- |
| Задание | Формула поиска | Ответ |
| 1. Найдите в тексте первые пять слов, состоящих из пяти букв | <[a-z]{5}> | Tired, could, would, worth, close |
| 2. Сколько в тексте шестибуквенных слов, начинающихся на букву s и заканчивающиеся на букву r? | <[Ss][a-z]{4}[Rr]> | 16 |
| 3. Найдите в тексте первые пять трёхбуквенных слов, начинающиеся на гласную букву | <[EeYyIiUuAaoO][a-z]{2}> | And, use, out, own, all |
| 4. Сколько в тексте слов, состоящих из двенадцати букв? По каким формальным признакам их можно сгруппировать? Приведите пример из каждой группы слов | <[a-z]{12}> | Ответ:31  Группы: |
| 5. Сколько в тексте слов с суффиксом -tion? Приведите пример использования такого слова в контексте | tion> | Ответ:59  Пример: to ask the question |
| 6. Есть ли в тексте слова, включающие четыре согласные буквы подряд? | |  | | --- | | [qwrtpsdfghjklzxcvbnmQWRTPSDFGHJKLZXCVBNM]{4} | | 24 |

**Уровень 3**

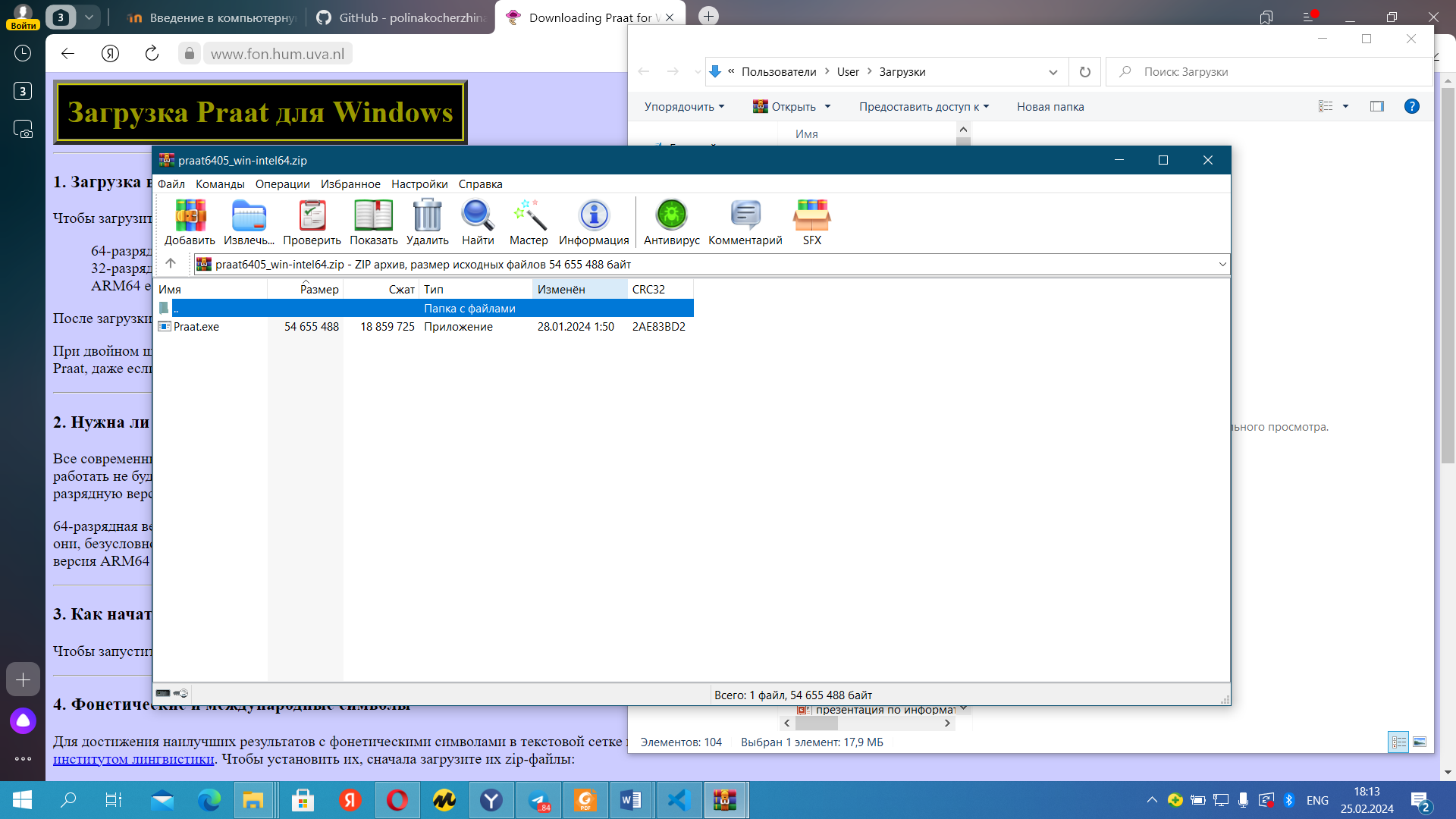
Из интернета я узнала, что **Praat** — это программа для анализа и реконструкции акустических речевых сигналов.

С её помощью можно:

* генерировать формы волн, широкие и узкие спектрограммы, контуры интенсивности и треки высоты тона;
* делать записи, редактировать записанный звук и извлекать отдельные звуки для дальнейшего анализа;
* получать информацию о высоте тона, интенсивности, формантах, импульсах и т. д.;
* усиливать определённые частотные регионы;
* сегментировать и маркировать слова, слоги или отдельные фонемы;
* представлять результаты в графической форме для печати.

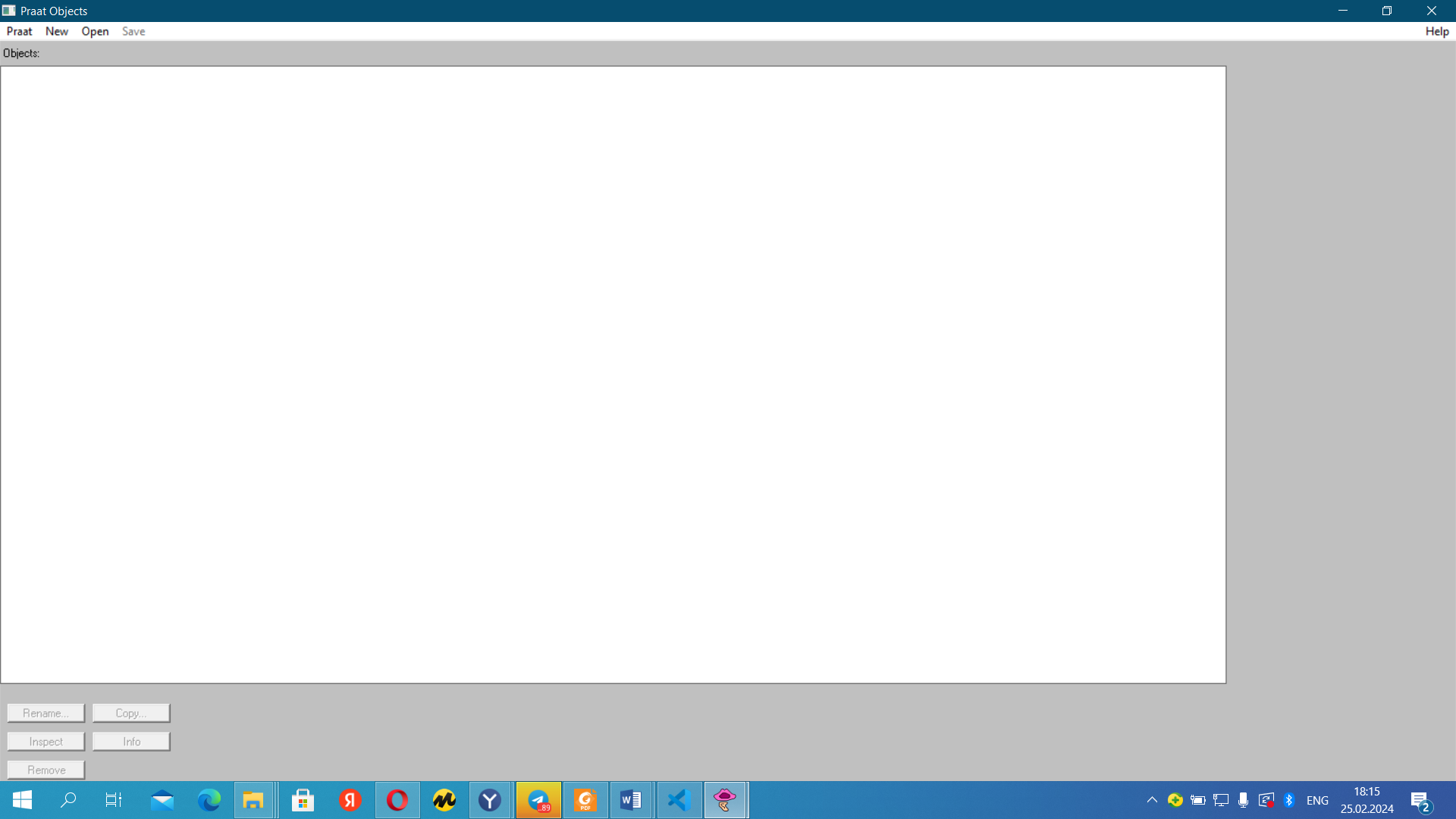
Перейдём к скачиванию самой программы Praat.



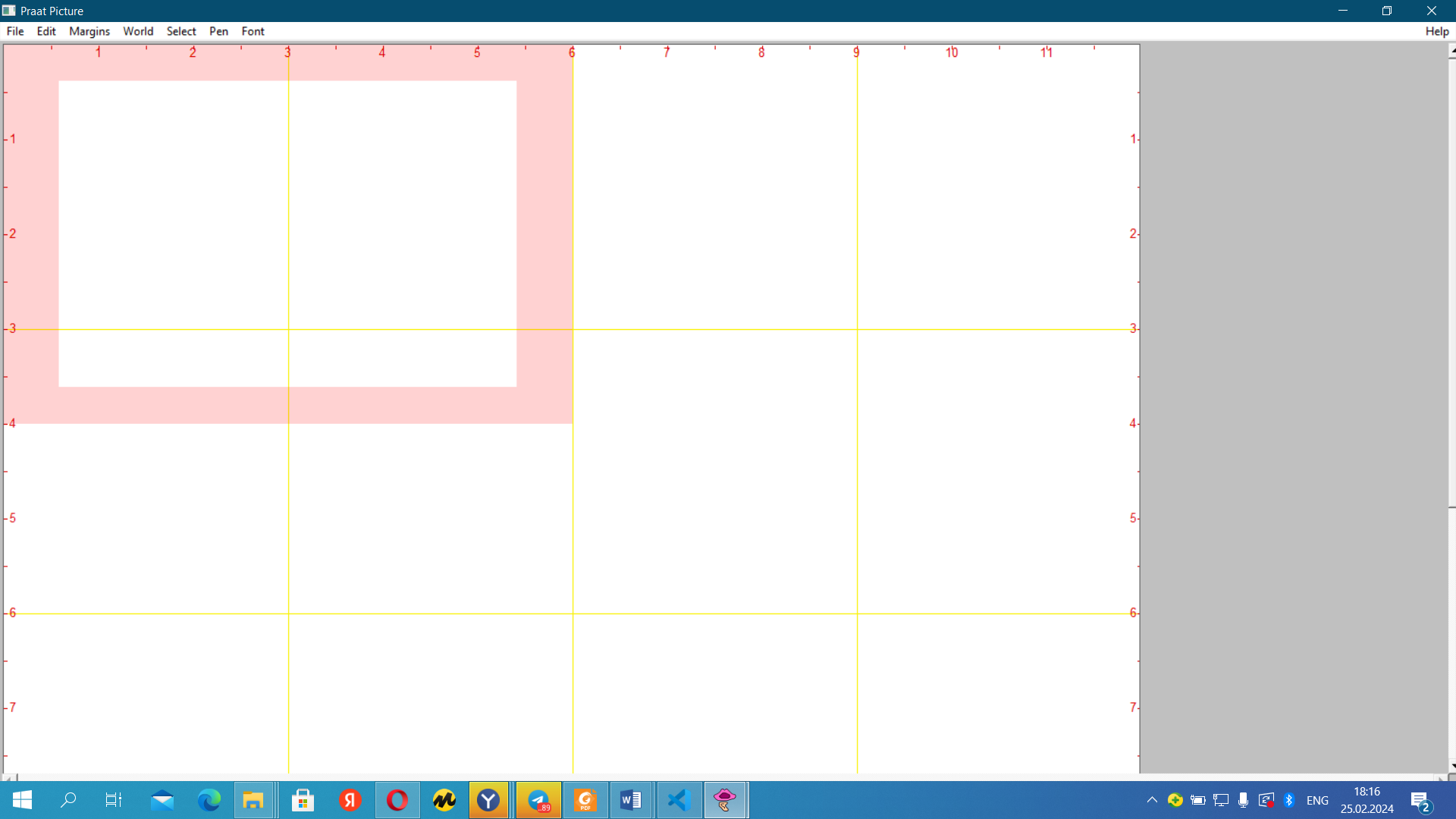


После открытия этой программы нам всплывает 2 окна.

В **окне Объектов Praat** мы можем открывать, создавать и сохранять файлы. Это меню можно использовать для открытия различных редакторов и запросов, которые нам понадобятся для работы со звуковыми файлами. После открытия программы в списке объектов Praat нет объектов. Следовательно, список пуст, а кнопки внизу отключены и показаны серым цветом. Когда мы помещаете звуковые файлы в список, меню и кнопки становятся динамичными, и они могут меняться (появляться, исчезать или отключаться) в зависимости от выбранных объектов.



**Окно изображений Praat**используется для создания и отображения изображений издательского качества и открывается по умолчанию при запуске программы.



Приложение Praat имеет множество преимуществ и плюсов, которые делают его популярным среди исследователей звука и речи. Основные плюсы:

1.Praat доступен бесплатно для скачивания и использования, что делает его доступным для широкого круга пользователей.

2.Praat предоставляет широкий спектр инструментов для анализа звука и речи, включая измерение акустических параметров, создание спектрограмм, анализ формантов и многое другое.

3.Пользователи могут настраивать параметры анализа и синтеза звука в соответствии с их потребностями и требованиями исследования.

4.Praat поддерживает множество форматов аудиофайлов, что упрощает работу с различными типами данных.

5.Разработчики постоянно обновляют и улучшают Praat, добавляя новые функции и исправляя ошибки.

6.Praat может использоваться в сочетании с другими программами и языками программирования для более сложных задач анализа данных.