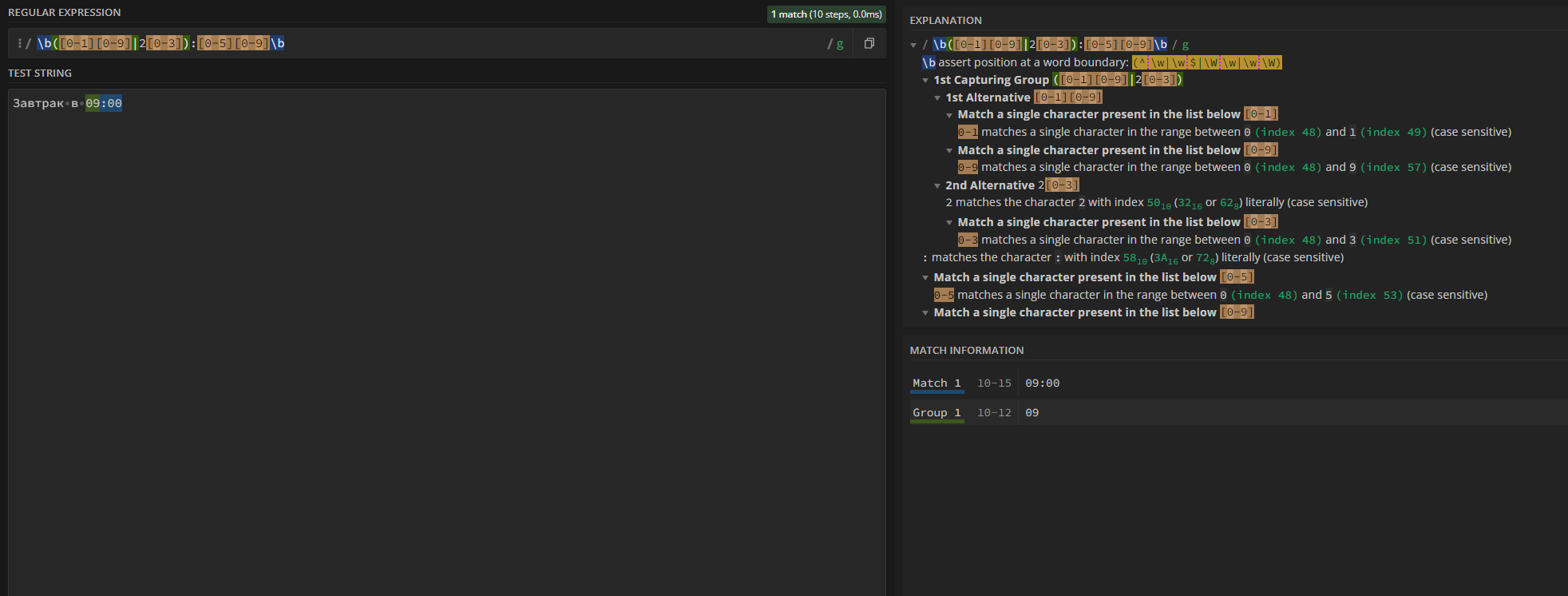
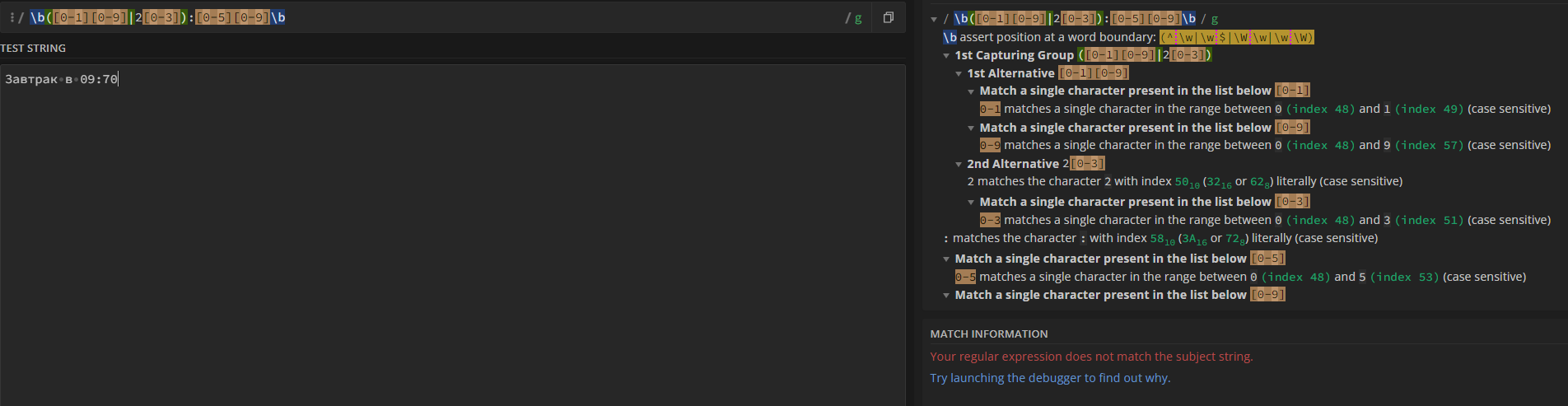
**Уровень 1**

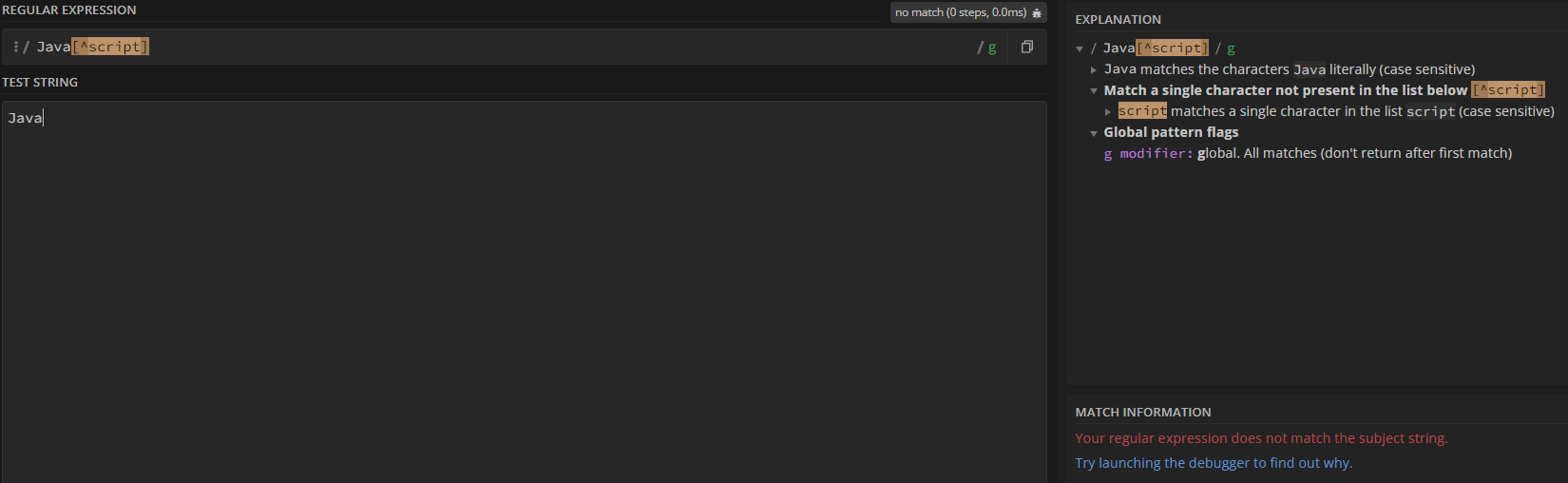
Найдём время:

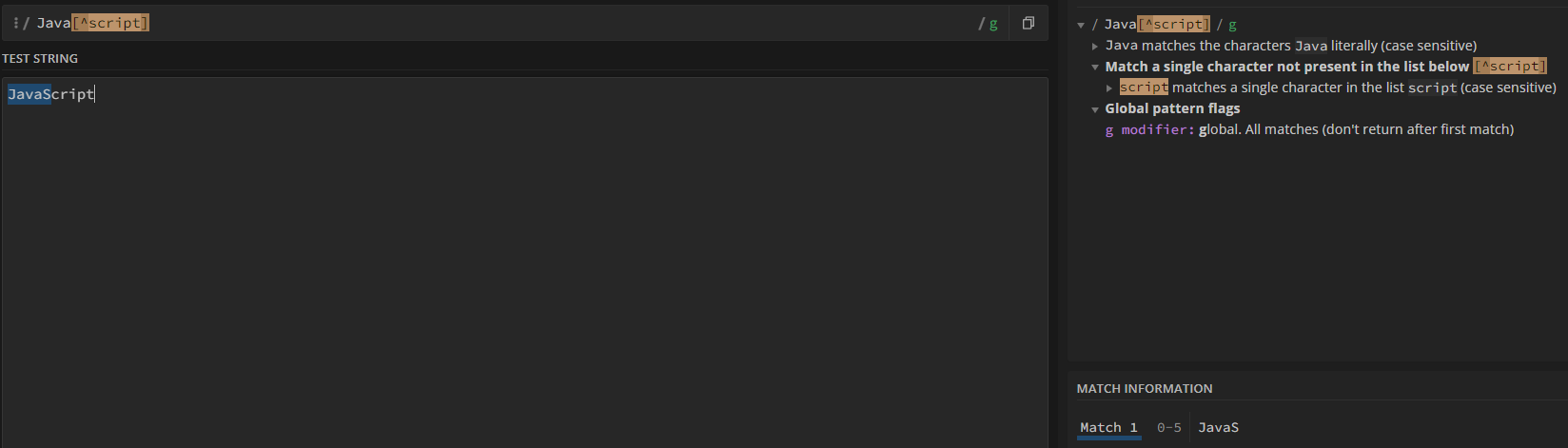


Проверим, как он себя ведёт при не верном запросе:



Видим, что результатов нет, а значит регулярное выражение составлено верно. Далее с Java:

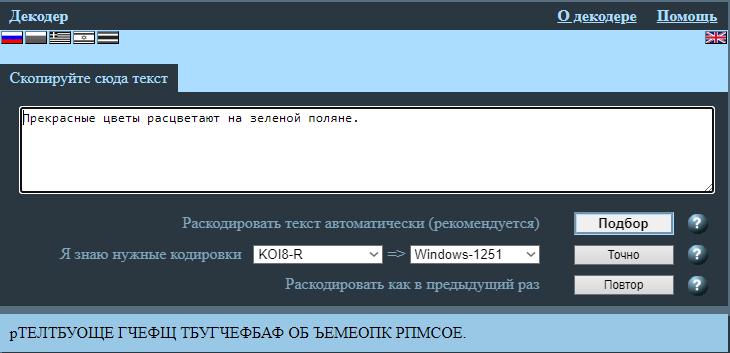




Видим, что в строке Java он не находит ничего, а в строке JavaScript уже находит.

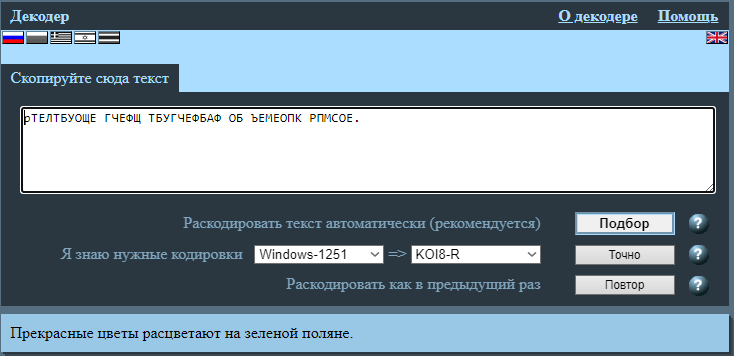
Затем напишу регулярное выражение для поиска HTML-цвета: #([A-Fa-f0-9]{6}), где # соответствует символу # в начале цветового кода в HTML, а выражение после решётки – это группа которая соответствует шести символам из шестнадцатеричных цифр и букв от A до F (независимо от регистра).

После запишу регулярное выражение, которое найдёт как всё арифметическое действие, так и (через группы) два операнда: ([-+]?\d+(\.\d+)?)\s\*([-+\*/])\s\*([-+]?\d+(\.\d+)?). В этом выражении ([-+]?\d+(\.\d+)?) – это группа, которая соответствует первому операнду в выражении. [-+]? позволяет наличие знака “+” или “-“ перед числом. \d+(\.\d+)? соответствует целочисленному или десятичному числу; \s\* - соответствует нулю или более пробельным символам, которые могут находиться между операндами и оператором; ([-+\*/]) - это группа, которая соответствует оператору “+” или “-“ или “\*” или “/”; \s\* - соответствует нулю или более пробельным символам, которые могут находиться между оператором и вторым операндом; ([-+]?\d+(\.\d+)?) - это группа, которая соответствует второму операнду в выражении, и она аналогична первому операнду.

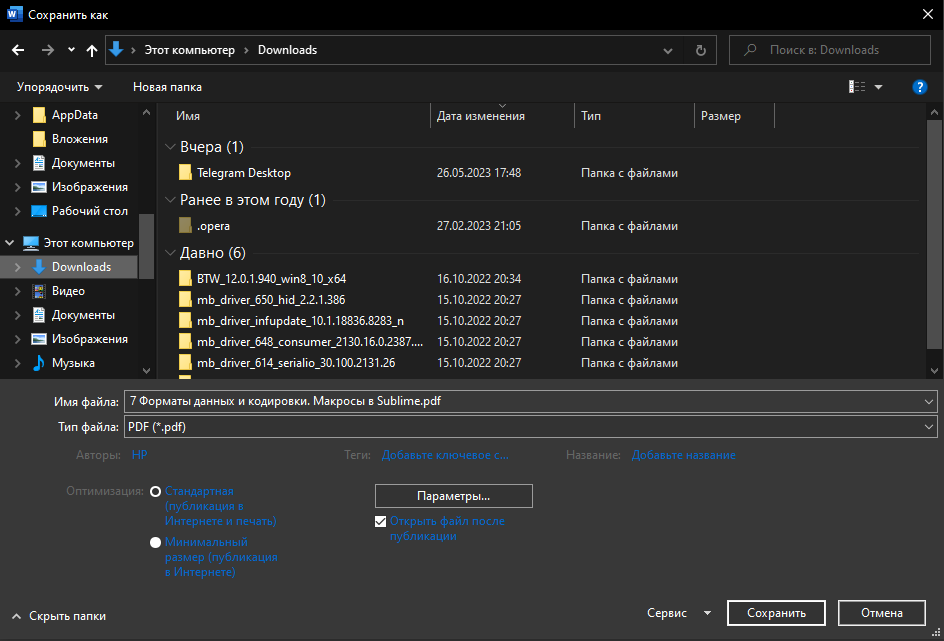
Далее попробуем закодировать данный текст: “ Прекрасные цветы расцветают на зеленой поляне.” 

Получим следующий результат: рТЕЛТБУОЩЕ ГЧЕФЩ ТБУГЧЕФБАФ ОБ ЪЕМЕОПК РПМСОЕ.

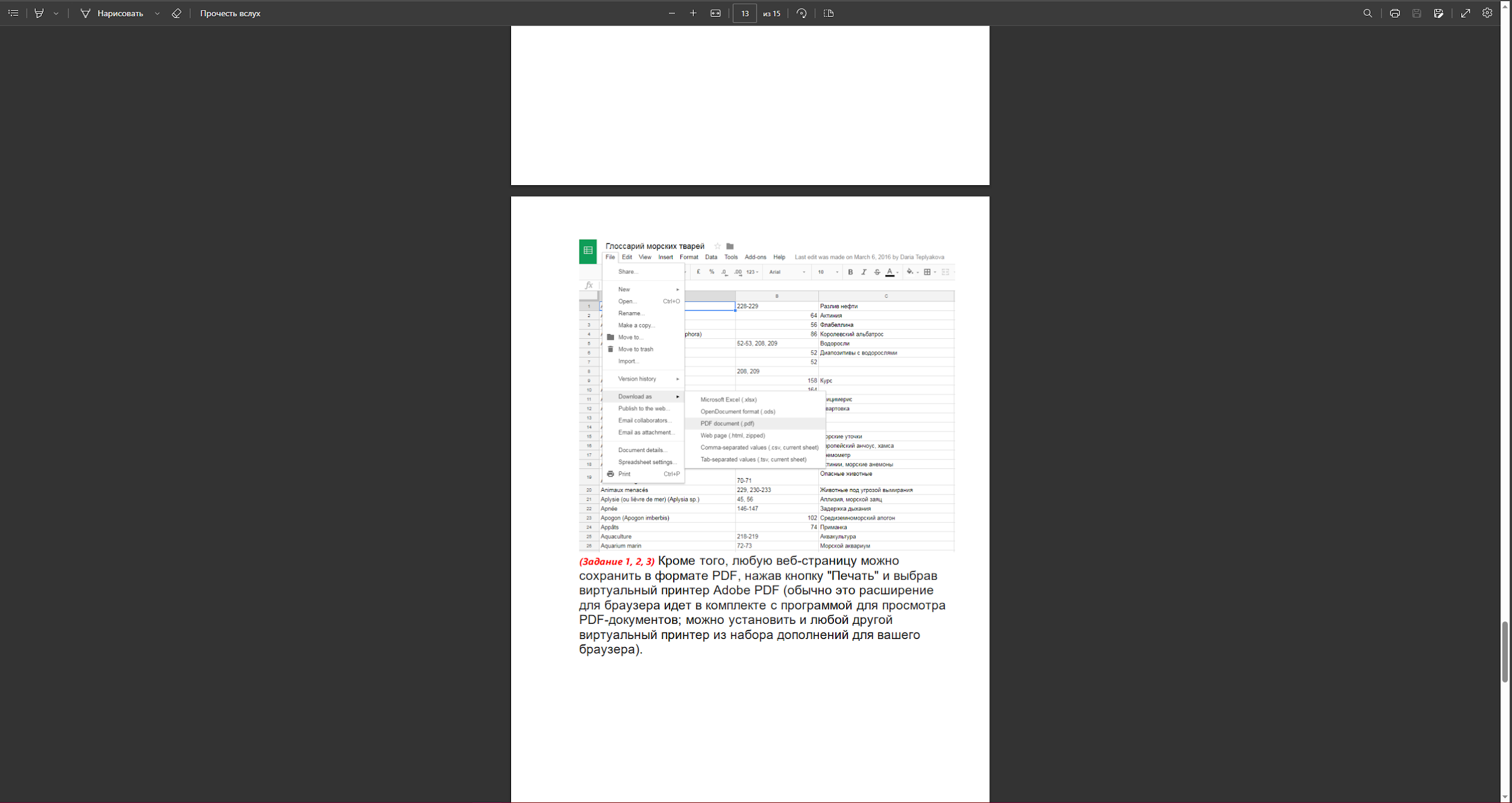
Декодировав обратно, получим:



Далее конвертируем файл типа .docx в .pdf:



Получим вполне сносный пдф:



Конспект всего того, что содержалось в файле:

Информация на цифровых носителях хранится в закодированном виде и представляет собой последовательность 0 и 1. Упорядоченность информации обеспечивается файловой системой, которая определяет формат содержимого и способ его хранения. Большие файлы записываются в кластеры, но при недостатке свободного места они могут быть разделены и записаны в разные места на диске, что называется фрагментацией. Дефрагментация позволяет перераспределить файлы для хранения их в непрерывной последовательности кластеров и улучшить скорость чтения. Расположение файла в такой структуре описывается с помощью пути. У файла есть расширение, которое указывает на его тип и помогает правильно раскодировать информацию. Например, при изменении расширения картинки с jpg на txt и ее открытии в текстовом редакторе, мы увидим непонятные символы, так как байтовая последовательность будет неправильно расшифрована. Ниже представлена небольшая история кодировок текста:

* ASCII (1963 год) — 7-битная кодировка, включающая 128 символов.
* ISO/IEC 646 (1972 год) — группа кодировок, основанных на ASCII.
* KOI8 (1974 год, СССР) — 8-битная кодировка, включающая латинские и русские символы.
* ISO/IEC 8859 (середина 80-х) — группа 8-битных кодировок, основанных на ASCII.
* Windows-125x (начало 90-х) — группа кодировок, разработанных для ОС Windows.
* UTF-8 (Unicode Transformation Format, 8-bit) — кодировка текста, позволяющая хранить символы Юникода.

Unicode - стандарт кодирования символов, включающий знаки почти всех письменных языков мира.

Кракозябры в тексте могут указывать на неправильную кодировку, которую нужно определить для правильного отображения. Формат PDF (Portable Document Format) обеспечивает одинаковое отображение документа на разных устройствах и операционных системах. Создание PDF-документов обычно осуществляется через экспорт из специализированных программ, чтобы сохранить форматирование и шрифты при просмотре на других компьютерах.

**Уровень 2**

Так как в самом файле я нашёл лишь теоретический материал и никакого конкретного задания, то решил просто законспектировать в сжатом достаточно формате представленную информацию.

Язык разметки - набор символов, вставляемых в текст для передачи информации о его выводе или строении. TEI (Text Encoding Initiative) - стандартизированный язык разметки для электронной публикации, представления документов и создания структурированных файлов. Состоит он из TEI Header (паспорт документа), FACSIMILE (копия источника) и текста. Структура заголовка представляет собой дерево, где теги вложены друг в друга. Теги закрываются в обратном порядке, т.е. тег, который был открыт первым, закрывается последним. Все открытые теги должны быть закрыты!

**<fileDesc>** (описание файла) содержит исчерпывающее библиографическое описание источника файла.

**<titleStmt>** группирует информацию о названии произведения и о тех, кто несет ответственность за его содержание.

**<title>** содержит название произведения, будь то статья, книга, журнал или серия, включая все альтернативные названия или подзаголовки.

**<respStmt>** содержит информацию о работе над текстом.

**<persName>** содержит собственное имя или название.

**<resp>** содержит информацию, описывающую тип персональной интеллектуальной ответственности за приводимые в тексте сведения.

**<sourceDesc>** содержит библиографическое описание экземпляра текста (или текстов), который служил оригиналом для создания или генерирования электронного варианта.

Текстовые редакторы, такие как Sublime и Notepad++, позволяют подсвечивать синтаксис языков разметки, таких как XML.

Анализ текста в pdf-файле включает чтение документа, осмысление его структуры и извлечение метаданных. Создается TEI документ с двумя частями: TEI-Header (содержит метаданные) и TEXT. Теги разметки в TEI HEADER включают информацию о документе и служебные данные. После разметки структуры документа можно проверить правильность XML-синтаксиса онлайн.

В документах нужно отмечать упоминания лиц, географических объектов и институций. Структурные элементы текста:

В классических западных документах фраза "Во имя Отца, и Сына, и Святого Духа" используется в качестве invocatio, а в русской практике обычно рисуют крест. Intitulatio указывает лицо, от которого исходит документ. Inscriptio указывает лицо, которому документ адресован. Dispositio содержит распоряжения и условия сделки, а corroboratio удостоверяет подлинность документа. Sanсtio указывает последствия для нарушителя, а datum содержит дату и место составления. Apprecatio, как благопожелание, редко используется в русской дипломатике.

Посмотреть т.н. "диффы" можно, установив плагин EasyDiff для Sublime или Compare для Notepad++, но проще всего воспользоваться каким-нибудь онлайн-сервисом -- [например, вот этим](https://www.diffchecker.com/).