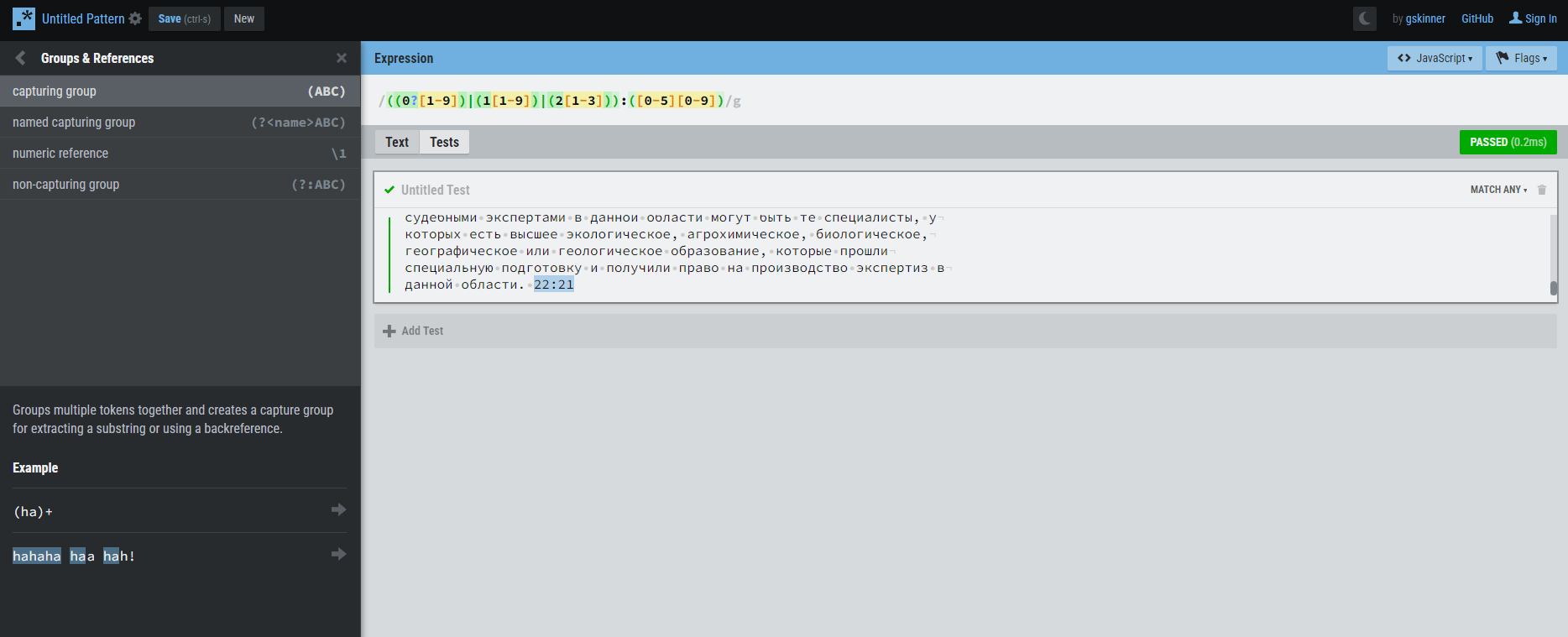
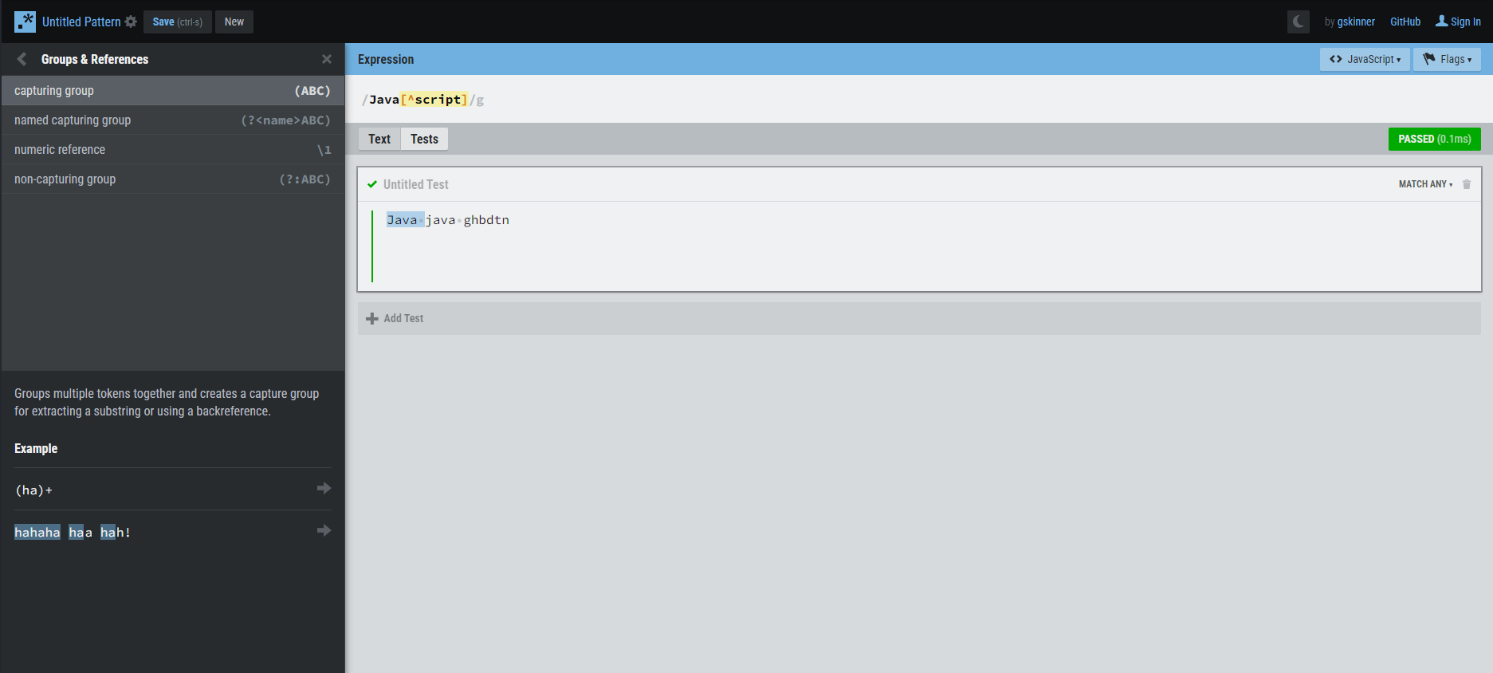
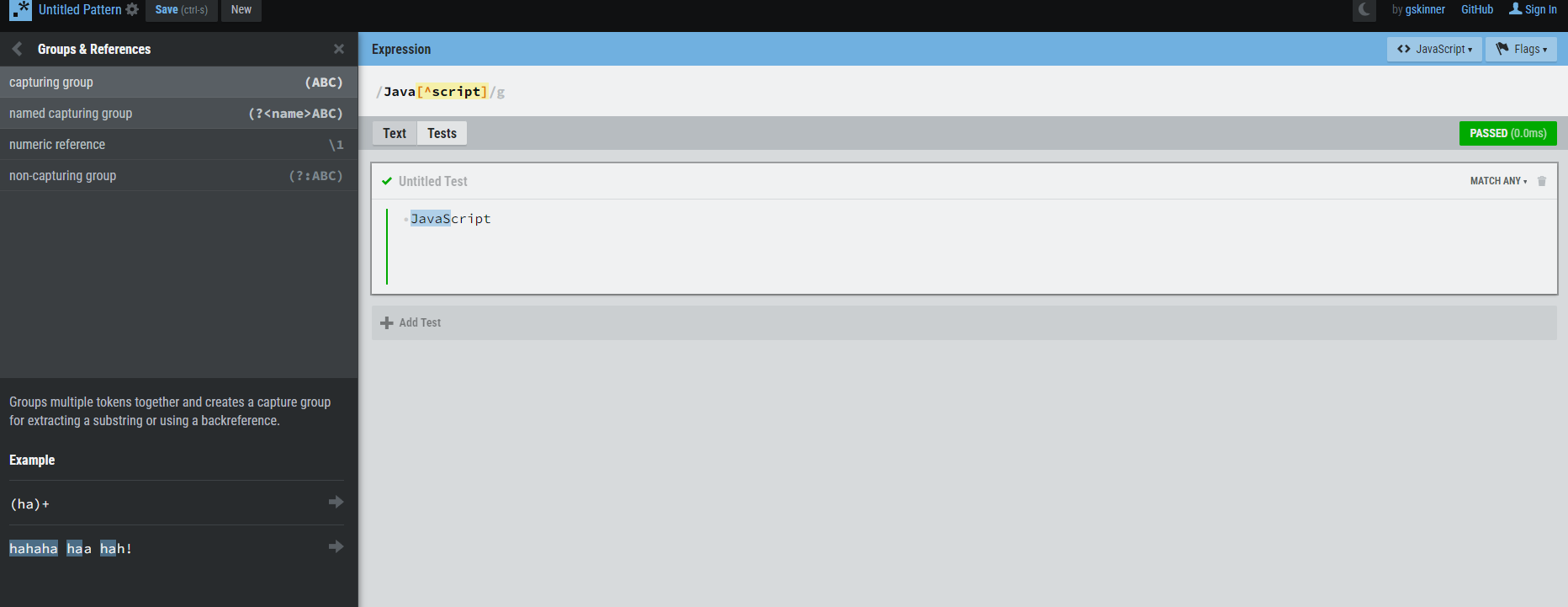
**Регулярные выражения**

Задание1

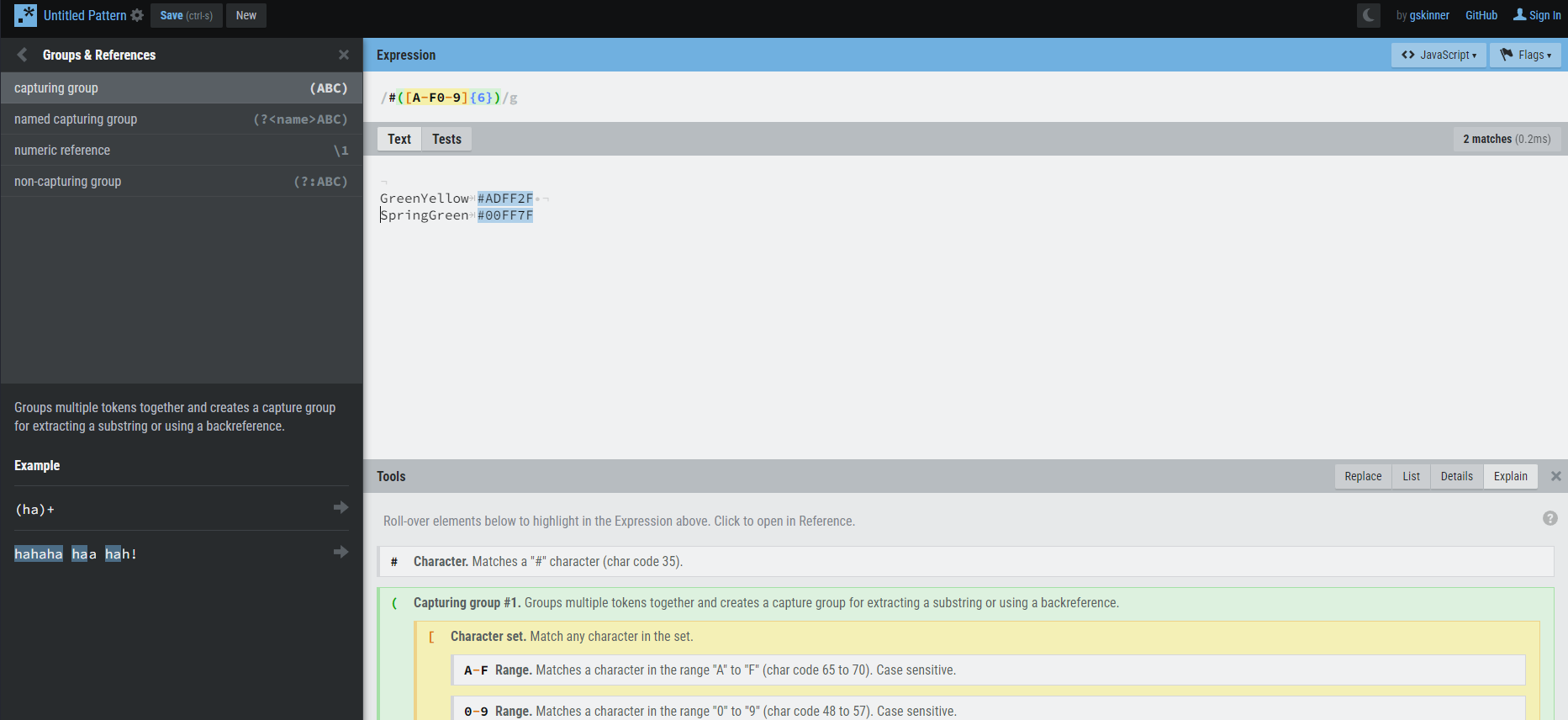


Задание2





Задание3



Задание4



**Форматы данных и макросы в Sublime**

Конспект

**Хранение информации на диске**

Информация на цифровых носителях хранится в **закодированном виде** и представляет собой **последовательность 0 и 1**. Как это устроено физически? В основе запоминающего устройства может лежать любой физический эффект, который обеспечивает приведение системы к двум устойчивым состояниям: одно из них будет обозначать 0, другое - 1. Так, в современных устройствах используются:

* наличие/отсутствие заряда в конденсаторе
* направление намагниченности
* отражение/рассеяние света от поверхности CD, DVD или Blu-ray диска

Минимальная единица хранения информации - 1 бит. С помощью него можно закодировать два значения - 0 и 1.

Упорядоченность информации обеспечивается **файловой системой**, которая определяет формат содержимого и способ его хранения. Иными словами, файловая система - это раздел жесткого диска, выделенный для хранения файлов на компьютере и другом электронном оборудовании: мобильных телефонах, фотоаппаратах и т.д. Рассмотрим ее устройство на примере FAT - файловой системы, впервые созданной в 1976-1977 годах Биллом Гейтсом и Марком МакДональдом. Ее модификации до сих пор используются, например, на флешках.

Обычно большие файлы записываются в кластеры, идущие подряд, но, когда свободного места на диске остается мало, нужного количества свободных ячеек подряд может не быть. Однако если суммарное количество свободных кластеров достаточно для записи файла, он записывается по кусочкам в разные места диска. Это называется **фрагментацией**. Правда, разбитые на кусочки файлы читаются гораздо медленнее, поэтому существует **дефрагментация** — перераспределение файлов на диске таким образом, чтобы каждый из них хранился в непрерывной последовательности кластеров.

**Форматы данных и кодировки**

Практически любая файловая система иерархична: файлы объединяются в каталоги (папки), которые образуют одно или несколько деревьев.

**Краткая история кодировок**

1. **ASCII** (1963 год) — 7-битная кодировка, включающая в себя 128 символов: 33 непечатных управляющих символа (влияющих на обработку текста и пробелов) и 95 печатных символов, включая цифры, буквы латинского алфавита в строчном и прописном вариантах и ряд пунктуационных символов. Недостаток: в ASCII не было букв и символов многих национальных алфавитов.
2. **ISO/IEC 646** (1972 год) — группы кодировок, основанных на ASCII, где редко используемые символы из ASCII заменялись на необходимые. Группа включала в себя варианты кодировки для Канады, Китая, Кубы, Германии, Дании, Финляндии, Франции, Великобритании, Греции, Венгрии, Ирландии, Японии, Южной Кореи, Мальты, Норвегии, Швеции и Югославии. Стандартизация кодировок группы позволила свести различия между кодировками разных стран к минимуму.
3. **KOI8** (1974 год, СССР) — Код Обмена Информацией. Как следует из названия, это была 8-битная кодировка, что позволяло включить в нее в два раза больше символов. KOI8 включала в себя цифры, буквы латинского и русского алфавита, а также знаки пунктуации, спецсимволы и псевдографику.
4. **ISO/IEC 8859** (середина 80-х). Ранние кодировки были ограничены 7 битами из-за особенностей некоторых протоколов передачи данных. Однако со временем эти ограничения свою актуальность потеряли, в то время как необходимость в дополнительных символах для языков, использующих латинский алфавит, только росла. Поэтому в середине 80-х началась работа над группой 8-битных кодировок, получившей название ISO/IEC 8859. Все кодировки этой группы были основаны на ASCII.
5. **Windows-125x** (начало 90-х) — группа кодировок, разработанная компанией Microsoft для ОС Windows. В Windows-1251 вошли все символы русского и близких к нему языков: украинского, белорусского, болгарского, сербского и македонского. На практике этого оказалось достаточно, чтобы кодировка Windows-1251 закрепилась в интернете вплоть до распространения UTF-8.
6. **UTF-8** (Unicode Transformation Format, 8-bit — «формат преобразования Юникода, 8-битный», 1993 год) — одна из общепринятых и стандартизированных кодировок текста, которая позволяет хранить символы Юникода, используя переменное количество байт (от 1 до 6). Полностью совместима с ASCII: все символы ASCII в UTF-8 кодируются 7 битами (первый - не значащий). Каждый символ кодировки, отличный от ASCII, состоит из ведущего байта, указывающего длину последовательности, и одного или нескольких продолжающих байт. Юникодные символы можно вводить с клавиатуры с помощью одновременного нажатия комбинации клавиш Alt+XXXX (например, Alt-1, Alt-12, Alt-4657; цифры набираются на Numpad'e) в Windows и с помощью комбинации клавиш Shift+Opt+XXXX на Mac OS.

**Unicode** – это стандарт кодирования символов, включающий в себя знаки почти всех письменных языков мира. В настоящее время стандарт является доминирующим в Интернете.

**Формат PDF**

**Portable Document Format (PDF)** — межплатформенный формат электронных документов, разработанный фирмой Adobe Systems. "Межплатформенный" в данном случае означает то, что документ будет одинаково отображаться на любом устройстве с любой операционной системой вне зависимости от того, где он был создан. Традиционным способом создания PDF-документов является виртуальный принтер, то есть документ как таковой готовится в своей специализированной программе — графическом или текстовом редакторе, а затем экспортируется ("публикуется") в формат PDF.

**Анализаторы текста**

1 Datavid Rover

## 2 SAS Text Miner

## 3 DiscoverText

## 4 IBM Watson

## 5 Google Cloud NLP

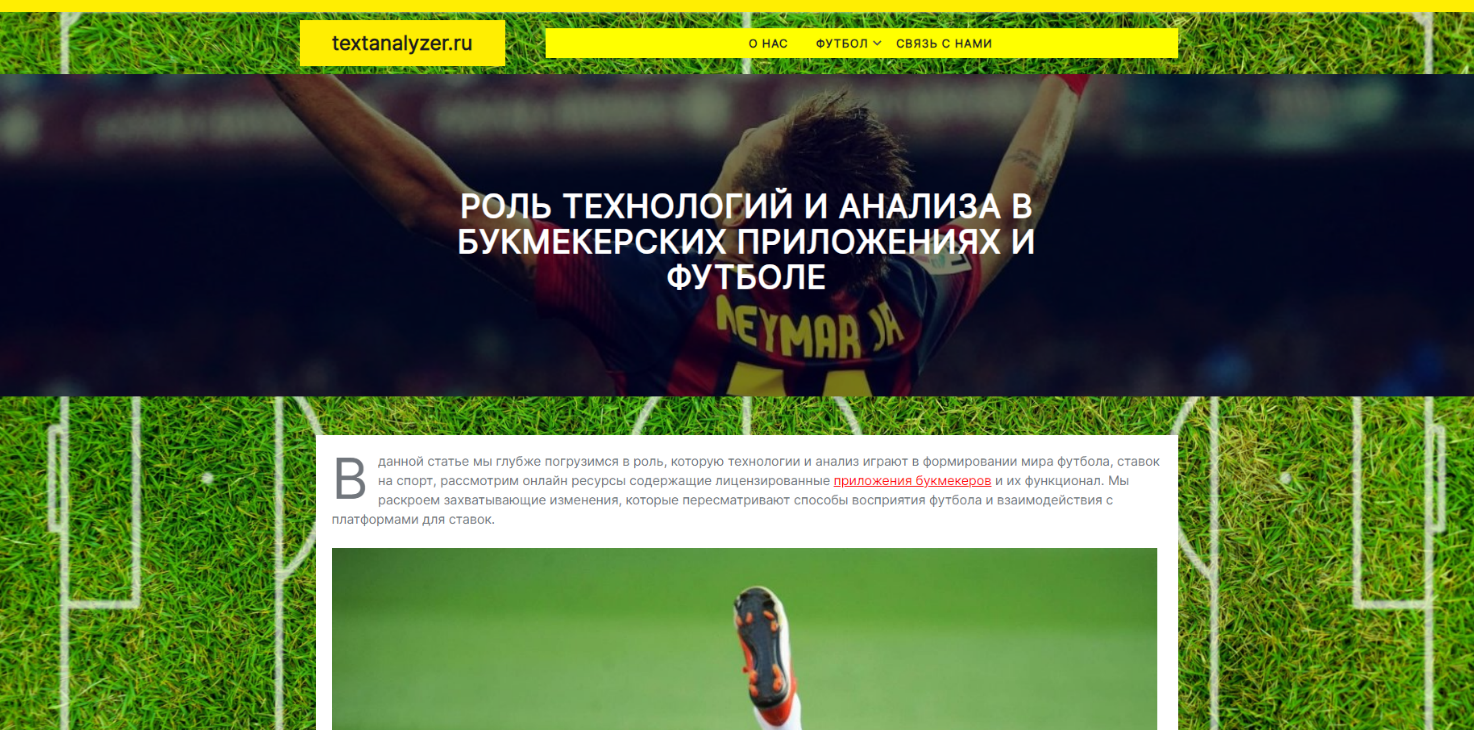
## 6 MeaningCloud

## 7 Textable

## 8 Amazon Comprehend

## 9 Aylien

## 10 Apache OpenNLP

****

анлакич((