

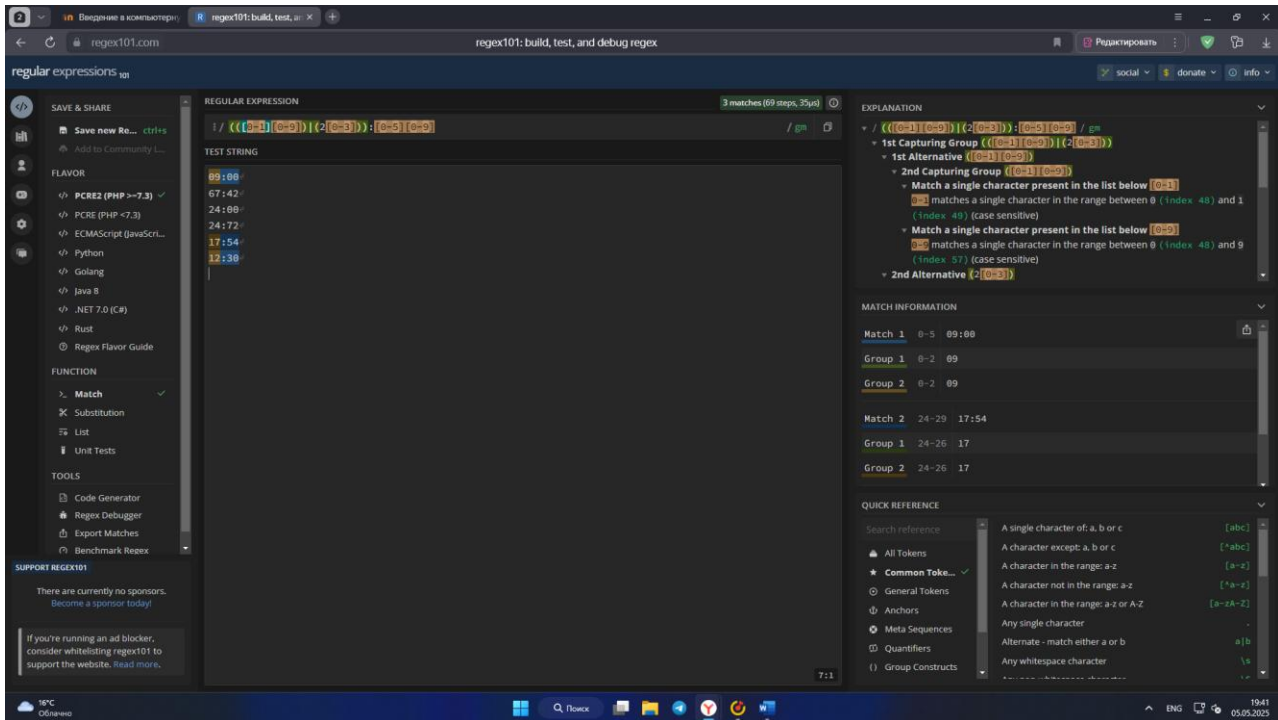
Лабораторная работа №12

Уровень 1

Для работы использовала сайт: <https://regex101.com/>

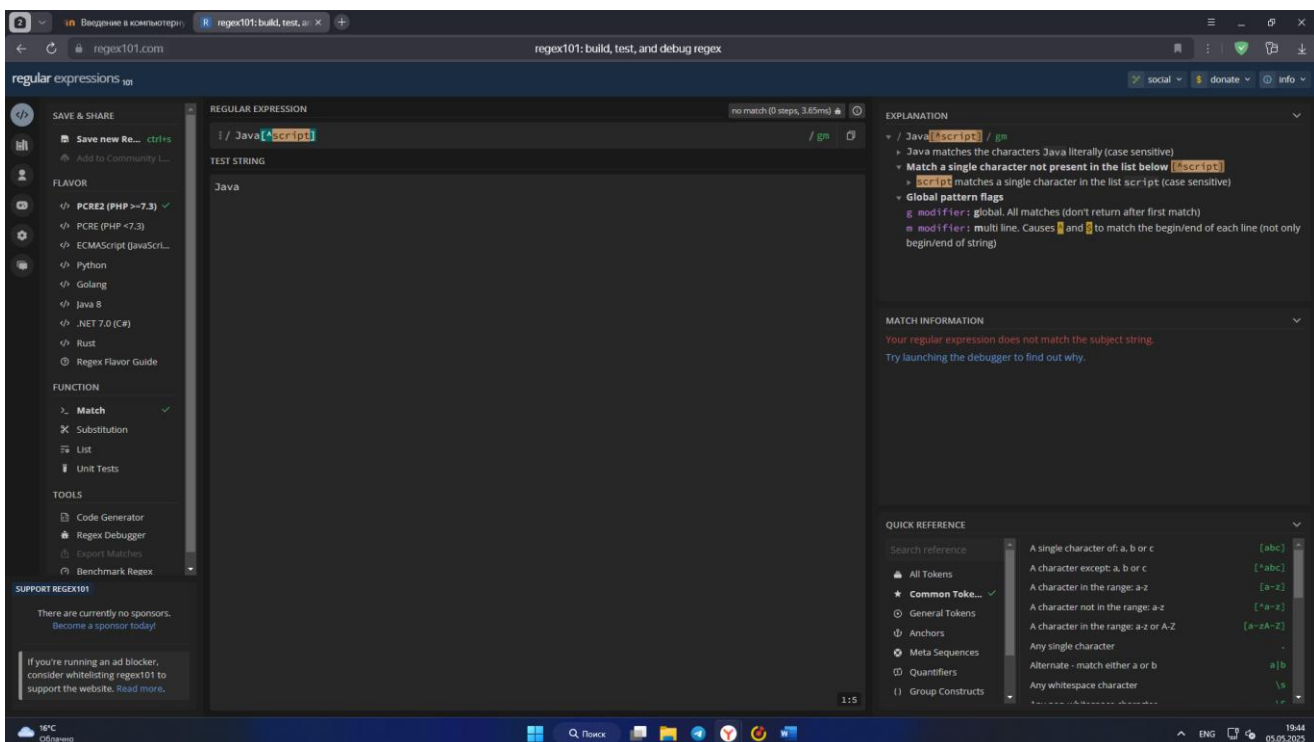
1.1. Найдите время

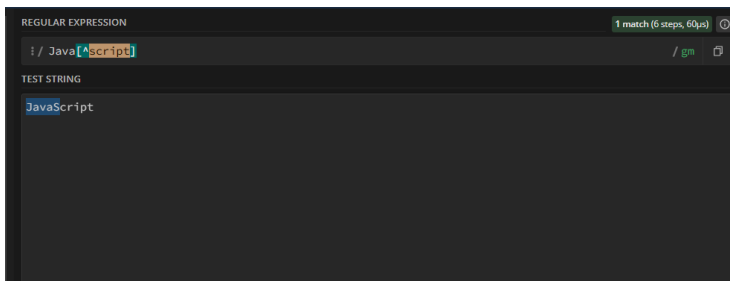
Время имеет формат часы:минуты. И часы, и минуты состоят из двух цифр, пример: 09:00. Напишите регулярное выражение для поиска времени в строке: "Завтрак в 09:00". Учтите, что "37:98" – некорректное время.



1.2. Java[[^]script]

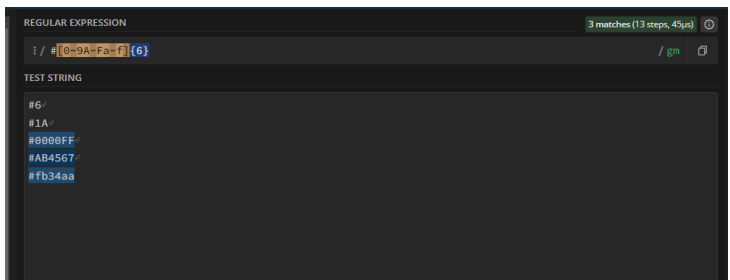
Найдет ли регулярка Java[[^]script] что-нибудь в строке Java? А в строке JavaScript?





1.3. Цвет

Напишите регулярное выражение для поиска HTML-цвета, заданного как #ABCDEF, то есть # и содержит затем 6 шестнадцатеричных символов.



1.4. Разобрать арифметическое выражение

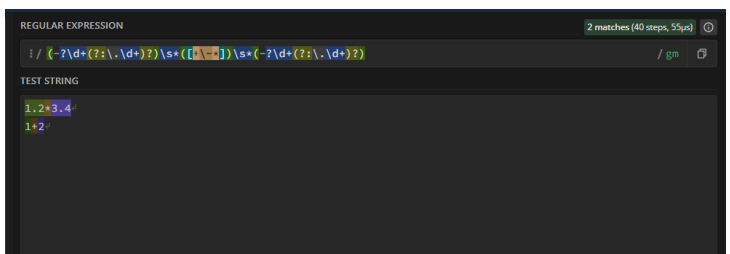
Арифметическое выражение состоит из двух чисел и операции между ними, например:

- 1 + 2
- 1.2 * 3.4
- -3 / -6
- -2 - 2

Список операций: "+", "-", "*", "/"

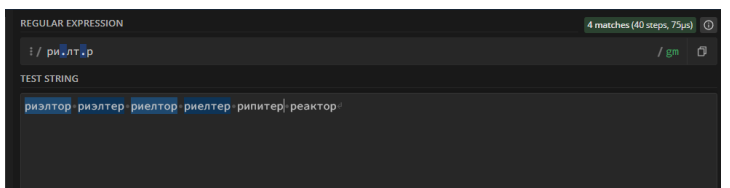
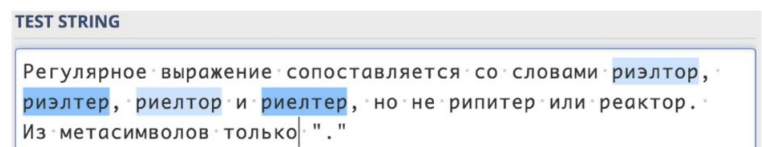
Также могут присутствовать пробелы вокруг оператора и чисел.

Напишите регулярное выражение, которое найдёт как всё арифметическое действие, так и (через группы) два операнда.



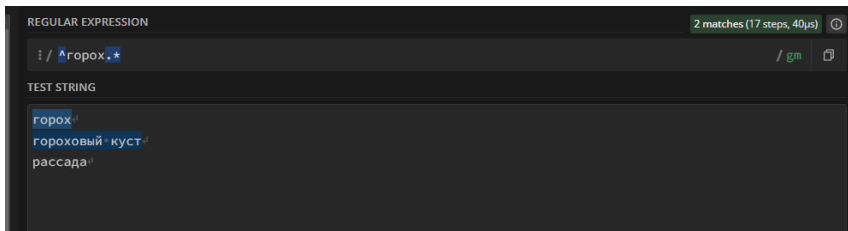
2.1

Напишите регулярное выражение, которое сопоставится со словами риэлтор, риэлтер, риелтор и риелтер (но НЕ сопоставится со словами рипитер и реактор, например), используя из метасимволов только . точку. То есть хочется такой результат:



2.2

Напишите регулярное выражение, которое полностью накроет любую строку, которая начинается на "горох":

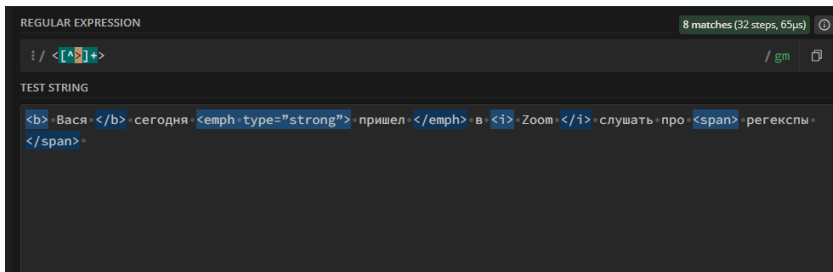


2.3

Напишите регулярное выражение, которое найдет ВСЕ теги (т.е. вида '<тег>' или '<тег атр="5">' или '</тег>') в строке:

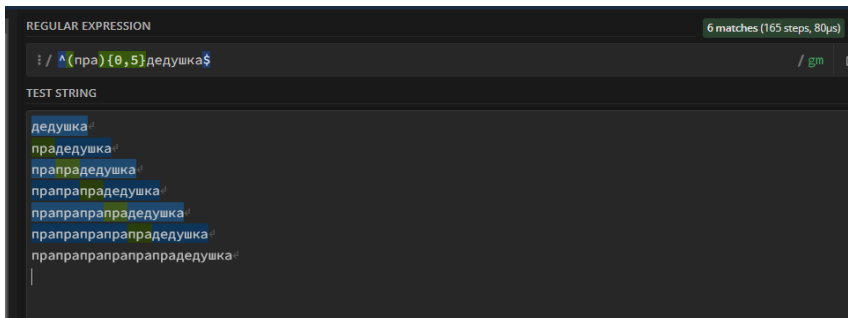
 Вася сегодня <emph type="strong"> пришел </emph> в <i> Zoom </i> слушать про регекспы

Внимание: выражение не должно съесть всю строку целиком! Надо именно найти все цепочки



2.4

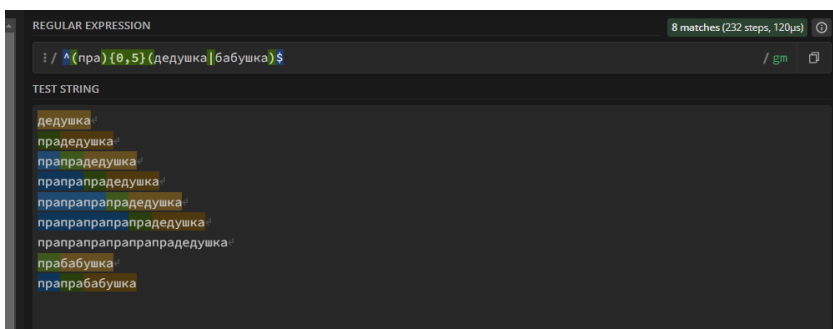
Напишите регулярное выражение, которое целиком накроет строки дедушка, прадедушка, прапрадедушка, прапрапрадедушка — и так до 5 пра, но не более 5



2.5

Напишите регулярное выражение, которое целиком накроет строки дедушка, прадедушка, прапрадедушка, прапрапрадедушка — и так до 5 пра, но не более 5.

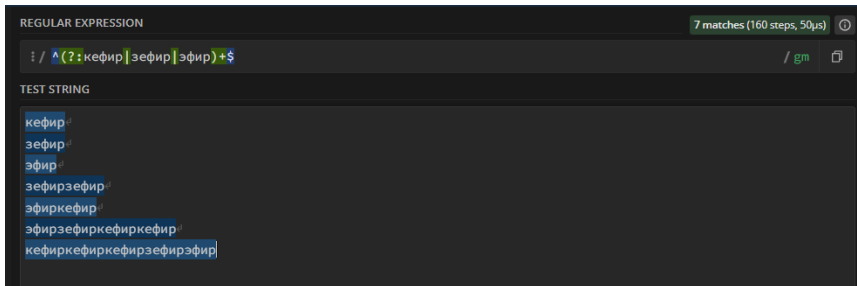
Добавьте бабушку в предыдущий регексп.



2.6

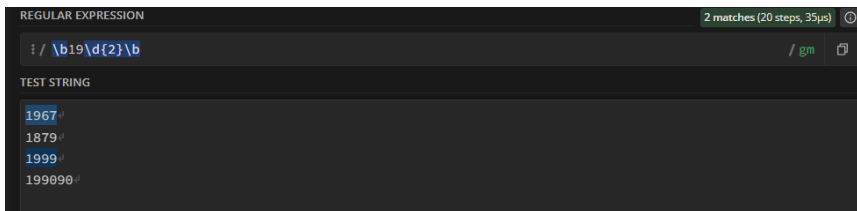
Напишите регулярное выражение, которое найдет все подстроки, в которых один или более раз повторяется 'кефир' или 'зефир' или 'эфир' (можно комбинировать слова). Т.е. выражение должно полностью сопоставляться с такими строками:

- кефир
- зефир
- эфир
- зефирзефир
- эфиркефир
- эфирзефиркефиркефир
- кефиркефиркефирзефирэфир



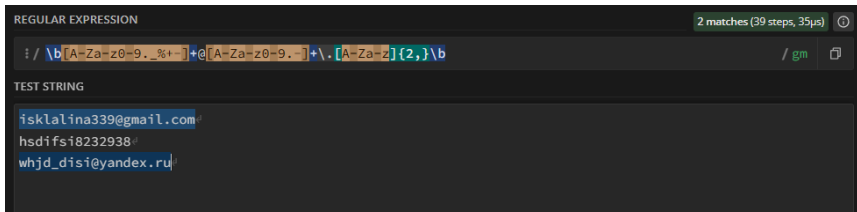
2.7

Напишите регулярное выражение, которое будет находить все года 20 века



2.8

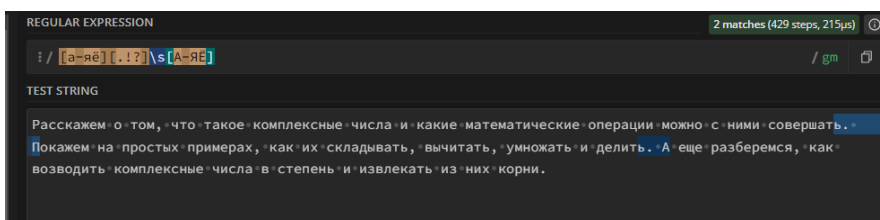
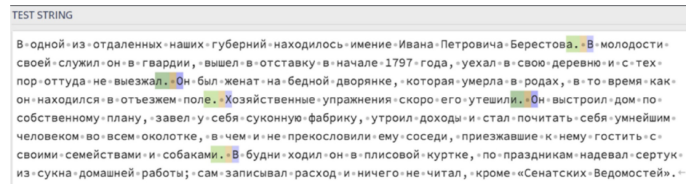
Напишите регулярное выражение, которое будет находить адреса электронной почты



2.9

Задание: сегментация на предложения (примитивная)

Напишите регулярное выражение, которое ловит такие случаи: после маленькой кириллической буквы стоит точка, восклицательный знак или вопросительный знак, потом пробел и большая кириллическая буква.



2.10

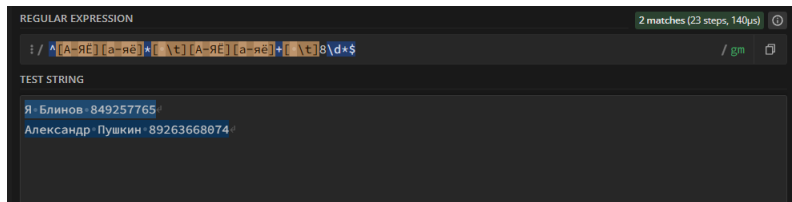
Задание: телефонный справочник

Напишите регулярное выражение, которое найдет такую комбинацию:

- одиночная кириллическая буква в верхнем регистре
- ноль или более кириллических букв в нижнем регистре
- пробел или таб
- одиночная кириллическая буква в верхнем регистре
- одна или более кириллических букв в нижнем регистре
- пробел или таб
- 8 и потом любое количество любых цифр

Я Блинов 849257765

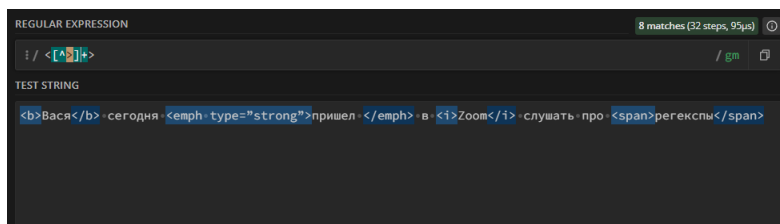
Александр Пушкин 89263668074

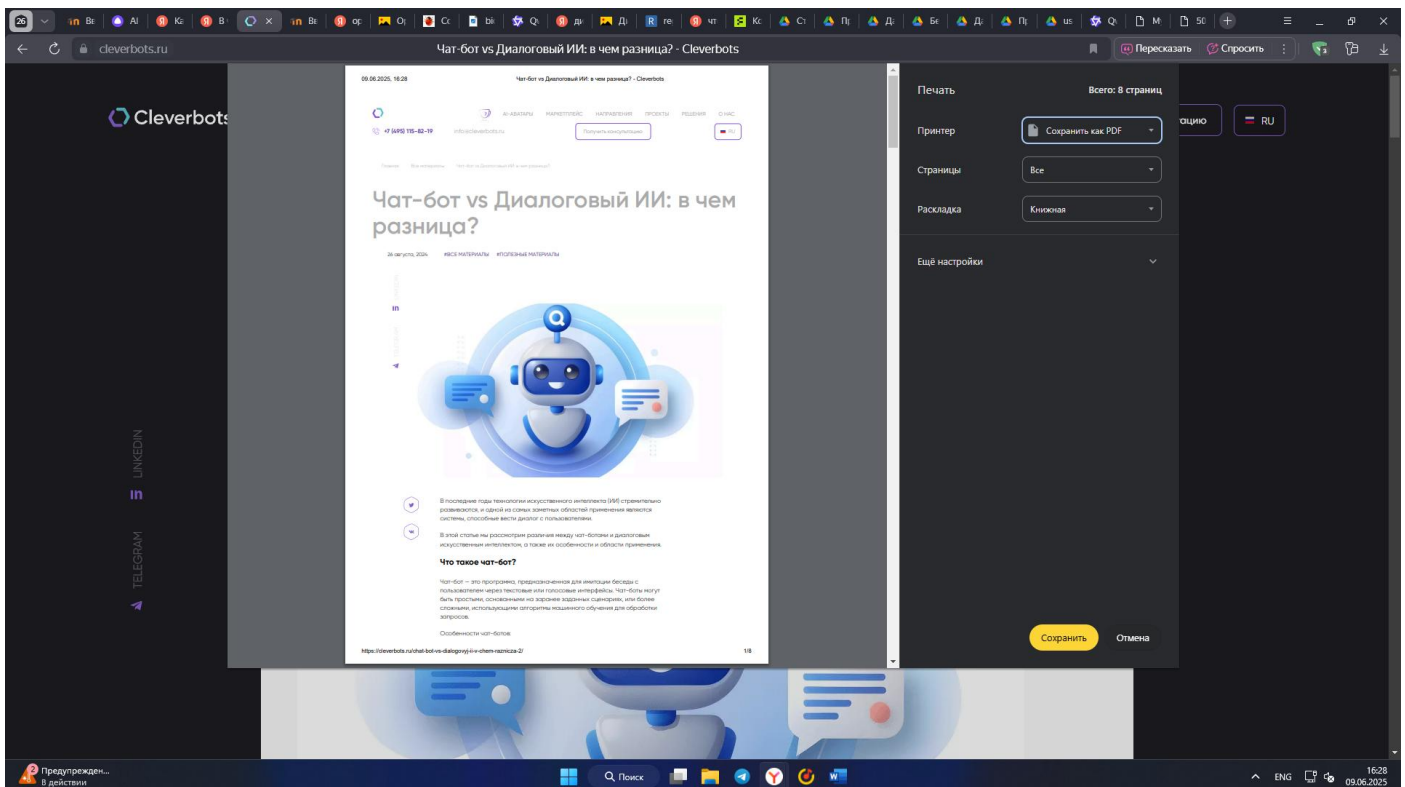
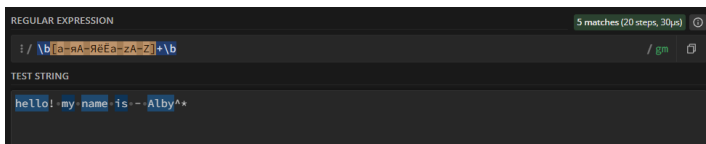


2.11

Напишите регулярное выражение, которое найдет ВСЕ теги (т.е. вида '<тег>' или '<тег атр="5">' или '</тег>') в строке

Вася сегодня <emph type="strong">пришел </emph> в <i>Zoom</i> слушать про регекспы





Форматы данных и кодировки.

Макросы в Sublime.

Хранение информации на диске

Информация на цифровых носителях хранится в закодированном виде и представляет собой последовательность 0 и 1.

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

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.

Файловая система - это раздел жесткого диска, выделенный для хранения файлов на компьютере и другом электронном оборудовании: мобильных телефонах, фотоаппаратах и т.д.

FAT - файловые системы впервые созданы в 1976-1977 годах Биллом Гейтсом и Марком МакДональдом. Ее модификации до сих пор используются, например, на флешках.

Название FAT расшифровывается как File Allocation Table, или "таблица размещения файлов". Это значит, что "адрес" каждого файла на диске прописан в специальной таблице. Дисковое пространство разбито на кластеры, и каждый файл занимает не менее одного из них.

Файловая система FAT состоит из следующих частей:

- Загрузочная запись
- FAT – та самая таблица соответствия файлов и кластеров, в которых они расположены
- Собственно файлы

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.

Расположение файла в такой структуре описывается с помощью пути. Например, hello.txt в строке C:\Users\Dereza\Documents\Library\hello.txt – это имя файла, а C:\Users\Dereza\Documents\Library - путь к нему.

Кроме того, у файла есть расширение - несколько символов после точки, которые обозначают его тип (формат): txt, doc, jpg, mp3, html... Расширение указывает на то, какая информация (текстовая, графическая и т.д.) находится в файле и как правильно ее декодировать.

Кодировка текста

Для перевода последовательности нулей и единиц в текстовые символы используется специальная таблица, которая называется кодировкой.

Основные кодировки:

- ASCII (7 бит, 128 символов) — только латиница и управляющие символы.
- KOI8 (8 бит) — поддерживает кириллицу.
- Windows-1251 — кириллица для Windows.
- UTF-8 (Unicode) — стандарт для всех языков (1–4 байта на символ).

Формат PDF

Преимущества:

- Сохраняет оформление на любых устройствах.
- Создается через виртуальный принтер (например, "Сохранить как PDF" в браузере).

Как создать PDF:

В Word/Google Docs: Файл → Сохранить как PDF.

Из браузера: Ctrl + P → Выбрать "Сохранить как PDF"