# Разбор домашнего задания

Роман Булгаков

Спикер курса

# Регулярные выражения: модуль re и его методы

Роман Булгаков

Спикер курса

# Регулярные выражения: шаблоны

Роман Булгаков

Спикер курса

# Задача «Слова на гласную»

**В**ходные данные:

text = 'AV is largest Analytics community of India'

- Выходные данные:
  - слова, которые начинаются на гласную букву

#### Правила для паттерна:

- Начало слова на одну из гласных букв: aeiou
- 2 Любой регистр: aeiou и AEIOU
- Нет ограничений на количество букв в слове

# Зачем нужны регулярные выражения?

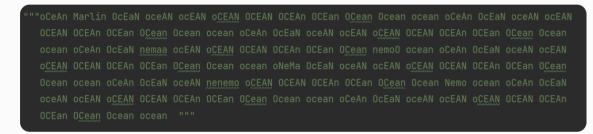
- Проверка формата данных (телефонный номер, email)
- Разбиение строки на подстроки;
   поиск, извлечение и замена символов
- Высокая скорость выполнения сложных операций по поиску и замене подстрок



Источник: habr.com

### Задача «В поисках Немо»

#### **В**ходные данные:



Возможна ошибка в окончании имени

#### Выходные данные:

• найти все упоминания Немо

#### Правила для паттерна:

- Нужны слова, которые начинаются на Nem или nem
- Окончание слова может быть разным (от 0 до 2 букв)

# Основы парсинга: модуль request и формат json

Роман Булгаков

Спикер курса

# Формат JSON

JSON (JavaScript Object Notation) — текстовый формат обмена данными. Является одним из стандартизированных форматов.

JSON представляет собой:

- набор пар ключей: значение (словарь, хеш-таблица и т. д.)
- упорядоченный набор значений (например, список)

```
"firstName": "Иван",
"lastName": "Иванов",
"address": {
    "streetAddress": "Московское ш., 101, кв.101",
    "city": "Ленинград",
    "postalCode": 101101
},
"phoneNumbers": [
    "812 123-1234",
    "916 123-4567"
]
```

Содержание типичного JSON-файла

# Итоги урока

- import requestsreq = requests.get('https://swapi.dev/api/people/3/')
- Формат JSON
- import json
- # десериализация JSONdata = json.load(file)data = json.loads(req.text)
- # сериализация JSONjson.dump(data, file)j\_data = json.loads(req.text)



# Модуль itertools

Роман Булгаков

Спикер курса

# Итоги модуля

- import re # работа с regular expressions
- my\_pattern = r'\b[aeiouAEIOU]\w+'result = re.findall(my\_pattern, text)
- import requests # запросы
   import json # работа с json форматом
- collections, itertools, datetime, decimal, matplotlib, numpy, pandas . . .

