

Техническое задание

Анализатор текста с использованием регулярных выражений

Проект: TextAnalyzer

Модуль: Анализ структурированных данных в тексте

Версия: 1.0

1. ОПИСАНИЕ ПРОЕКТА

Разработать консольное приложение для автоматического анализа текстовых данных и извлечения структурированной информации с использованием регулярных выражений

2. ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

2.1. Модуль извлечения email-адресов

- **Назначение:** Автоматическое извлечение email-адресов из произвольного текста
- **Требования к формату:**
 - Локальная часть: буквы (a-z, A-Z), цифры (0-9), символы . _ % + -
 - Доменная часть: буквы, цифры, точки, дефисы
 - Доменная зона: 2+ буквы
- **Ожидаемая точность:** >95%
- **Примеры валидных значений:**
 - user.name@domain.com
 - john_doe123@company.co.uk
 - support@sub-domain.org

2.2. Модуль парсинга дат

- **Назначение:** Выявление дат в формате ДД.ММ.ГГГГ
- **Требования к валидации:**
 - День: 01-31
 - Месяц: 01-12
 - Год: 1900-2099

- **Критерии приемки:** Не должны извлекаться некорректные даты (45.13.2025)
- **Формат вывода:** ДД.ММ.ГГГГ

2.3. Модуль идентификации именованных сущностей

- **Назначение:** Обнаружение слов, начинающихся с заглавной буквы
- **Область поиска:**
 - Русский алфавит: А-Я
 - Английский алфавит: A-Z
 - Поддержка составных конструкций через дефис
- **Исключения:** Не учитывать слова в начале предложения

2.4. Модуль статистики числовых данных

- **Назначение:** Подсчет общего количества цифровых символов в тексте
 - **Метрика:** Общее количество цифр (0-9)
 - **Точность:** 100%
-

3. ТЕХНИЧЕСКИЕ ТРЕБОВАНИЯ

3.1. Интерфейс системы

```
public class TextAnalyzer {  
  
    public static void findEmails(String text)  
  
    public static void findDates(String text)  
  
    public static void findCapitalizedWords(String text)  
  
    public static void countDigits(String text)  
  
}
```

3.2. Требования к выводу

- Четкая визуализация результатов по каждому модулю
- Нумерованный список найденных элементов

- Статистические данные в читаемом формате
- Обработка случаев отсутствия результатов

3.3. Производительность

- Обработка текстов до 10,000 символов
-

4. ТЕСТ-КЕЙСЫ ДЛЯ ПРИЕМКИ

Тестовый сценарий 1: Комплексный анализ

Входные данные:

Компания "Рога и копыта" основана 15.03.2005.

Контакты: director@roga.ru, support@roga-kopita.com.

Бюджет проекта: 1500000 рублей. Штаб-квартира в Москве.

Ожидаемые результаты:

Email-адреса:

1. director@roga.ru
2. support@roga-kopita.com

Даты:

1. 15.03.2005

Слова с заглавной буквы:

1. Компания
2. Рога
3. Москве

Общее количество цифр: 11

Тестовый сценарий 2: Границочные случаи

Входные данные:

Некорректные данные: 45.13.2025, user@.com, 999.999.9999

5. КРИТЕРИИ ПРИЕМКИ

Обязательные требования:

- Все 4 модуля реализованы и функционируют
- Регулярные выражения соответствуют спецификациям
- Корректная обработка граничных случаев
- Читаемый и структурированный вывод результатов