

БОЛЬШОЕ ЗАДАНИЕ ПО АНД

Вариант 7. «Окрашенность»

Подготовили: студентки
324 и 325 групп
Китова Е. Д.
Проскурня А.М.
Филимонова А.О.

Постановка задачи

Составить программу, оценивающую употребление в заданном русскоязычном тексте слов-названий различных цветов (белый, синий, бордовый, пурпурный, болотный и т.п.). Программа должна

- строить словарь разных цветов, встретившихся в тексте, и встречаются ли в нем сложные цвета (светло-голубой, лимонно-желтый, лазурно-синий и др.);
- вычислять абсолютную и относительную частоту употребления каждого слова-названия цвета, наиболее часто встречающийся цвет и т.п.;
- выводить подсчитанную статистику в удобной, обозримой форме.

Какие задачи дополнительно решались? (Екатерина)

- Построение графиков частоты для каждого текста
- Поддержка автоматического распознавание сложных цветов, которых нет в словаре
- Построение сравнительных графиков по нескольким текстам
- Построение таймлайн-графика и таблицы появления цветов относительно повествования текста

Какие задачи дополнительно решались? (Аня П.)

- Поддержка автоматического распознавания сложных цветов любой длины, отсутствующих в словаре
- Поддержка интерфейса, позволяющего анализировать как приложенные, так и пользовательские тексты
- Анализ хроматического профиля для текстов по различным тематикам

Какие задачи дополнительно решались? (Аня Ф.)

- Поддержка автоматического распознавание простых и сложных цветов, которых нет в словаре
- Построение сравнительной таблицы для нескольких текстов
- Поддержка пользовательского интерфейса с несколькими режимами работы

Библиотеки (Екатерина)

- `natasha` - для морфологического анализа русского языка (лемматизация, определение частей речи)
- `nltk` - для стемминга (приведения слов к основе)
- `pandas` - для формирования сравнительных таблиц и экспорт в CSV
- `matplotlib` - построение графиков
- `re` - для работы с регулярными выражениями при предобработке текста
- `collections` - для подсчета частотности слов
- `pathlib` - работа с файлами и папками

Библиотеки (Аня П.)

- Nltk - токенизация, частота
- Py morphology2 - морфологический анализ
- Counter - подсчёт

Библиотеки (Аня Ф.)

- os - работа с файловой системой
- rutmorphy2 - морфологический анализ, токенизация и лемматизация
СЛОВ

Что понимали под простыми и сложными цветами?

Простой

- Отдельные слова, обозначающие один цвет
- Без дефисов и дополнительных частей
- Всегда присутствуют в словаре

Сложный

- Составные прилагательные, содержащие дефис
- Одна или более цветовых основ/простых лемм

СЛОЖНЫЙ ЦВЕТ

Словарь

- Составные цвета, внесённые в словарь
- В готовом виде: светло-голубой, темно-синий
- Если программа встречает их в тексте, они распознаются напрямую как известные цвета

Авто

- Конструкции с двумя и более дефисами, которых нет в словаре, но последняя часть является простым цветом из словаря.
- Программа определяет такие слова как цвета автоматически.

Структура (Екатерина)

- Программа состоит из одного основного файла `color_analyzer.py`, где реализован класс `ColorAnalyzer`.
- Каждая стадия анализа реализована отдельным методом(функцией) — от загрузки словаря до построения графиков.
- Взаимодействие с пользователем происходит через консольный интерфейс в трёх режимах: анализ одного текста, сравнение нескольких текстов и построение цветовой дорожки.

Структура (Аня П.)

- Интерактивный выбор текста для анализа
- Анализ текста: токенизация, нормализация и подсчёт статистик.
- Вывод полученных результатов
- Вывод результатов прикладной задачи по запросу пользователя

Структура (Аня Ф.)

- `color_dictionary.py` - работа со словарями цветов
- `colors_analyze.py` - главный исполняемый файл с классом `ColorAnalyzer`. Методы `ColorAnalyzer` загружают и обрабатывают текст, выделяют цвета из текста, а также подсчитывают статистику
- `comparsion.py` - сравнительный анализ
- `report.py` - формирование отчетов
- Взаимодействие с пользователем происходит через консольный интерфейс в двух режимах: анализ одного текста или сравнительный анализ нескольких текстов

Тест 1

Светло-голубые глаза, темно-коричневая шерсть, красная помада, нежно-розовые лепестки и ярко-желтое солнце.

Серо-буро-малиновый шарф, красно-серо-оранжевый свитер, беловато-синевато-желтоватый цвет.

Космически-черный автомобиль. в черной-черной комнате...

Тест 1

ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Всего слов: 22
Всего упоминаний цветов: 10
Плотность цветов относительно всех слов в тексте: 0.4545

СТАТИСТИКА ПО ТИПАМ ЦВЕТОВ:

Уникальных цветов всего: 10
Наличие сложных цветов: Да
Простых: 1
Сложных: 9
Соотношение уникальных: простых 10.0% / сложных 90.0%

ВСЕ НАЙДЕННЫЕ ЦВЕТА С ЧАСТОТАМИ: (всего 10)

Цвет	Абсолютная частота	Относительная частота(отн. всех слов)
беловато-синевато-желтоватый	1 раз(а)	0.045455
космически-черный	1 раз(а)	0.045455
красно-серо-оранжевый	1 раз(а)	0.045455
красный	1 раз(а)	0.045455
нежно-розовый	1 раз(а)	0.045455
светло-голубой	1 раз(а)	0.045455
серо-буро-малиновый	1 раз(а)	0.045455
темно-коричневый	1 раз(а)	0.045455
черный-черный	1 раз(а)	0.045455
ярко-желтый	1 раз(а)	0.045455

ТОП САМЫХ ЧАСТЫХ ЦВЕТОВ:

1. 'красный': 1 раз (4.5455%)
2. 'светло-голубой': 1 раз (4.5455%)
3. 'темно-коричневый': 1 раз (4.5455%)
4. 'нежно-розовый': 1 раз (4.5455%)
5. 'ярко-желтый': 1 раз (4.5455%)
6. 'серо-буро-малиновый': 1 раз (4.5455%)
7. 'красно-серо-оранжевый': 1 раз (4.5455%)
8. 'беловато-синевато-желтоватый': 1 раз (4.5455%)
9. 'космически-черный': 1 раз (4.5455%)
10. 'черный-черный': 1 раз (4.5455%)

АНАЛИЗ ТЕКСТА: Пользовательский файл: tests/test2 (2).txt

ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Общее количество слов: 22
Количество уникальных словоформ: 22

Общее количество слов-цветов: 10

Из них:

- простых цветов: 1
- сложных цветов: 9

Число разнообразных цветов: 10

СТАТИСТИКА ПО ЦВЕТАМ:

светло-голубой	Абс. частота: 1	Отн. частота: 0.0455
темно-коричневый	Абс. частота: 1	Отн. частота: 0.0455
красный	Абс. частота: 1	Отн. частота: 0.0455
нежно-розовый	Абс. частота: 1	Отн. частота: 0.0455
ярко-жёлтый	Абс. частота: 1	Отн. частота: 0.0455
серо-буро-малиновый	Абс. частота: 1	Отн. частота: 0.0455
красно-серо-оранжевый	Абс. частота: 1	Отн. частота: 0.0455
беловато-синевато-желтоватый	Абс. частота: 1	Отн. частота: 0.0455
космически-чёрный	Абс. частота: 1	Отн. частота: 0.0455
черной-чёрный	Абс. частота: 1	Отн. частота: 0.0455

Самые часто встречающиеся цвета:

- светло-голубой (число упоминаний: 1)
- темно-коричневый (число упоминаний: 1)
- красный (число упоминаний: 1)
- нежно-розовый (число упоминаний: 1)
- ярко-жёлтый (число упоминаний: 1)
- серо-буро-малиновый (число упоминаний: 1)
- красно-серо-оранжевый (число упоминаний: 1)
- беловато-синевато-желтоватый (число упоминаний: 1)
- космически-чёрный (число упоминаний: 1)
- черной-чёрный (число упоминаний: 1)

АНАЛИЗ ТЕКСТА: tests/test2_1.txt

ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Всего слов: 22
Уникальных слов(лемм): 22
Общее упоминание цветов: 10
Уникальных цветов: 10
Плотность цветов: 0.4545

СТАТИСТИКА ПО ЦВЕТАМ:

Самые частые цвета: беловато-синевато-желтоватый, космически-черный, красно-серо-

СЛОЖНЫЕ ЦВЕТА (9):

- беловато-синевато-желтоватый (авто): 1 упоминаний
- космически-черный (авто): 1 упоминаний
- красно-серо-оранжевый (авто): 1 упоминаний
- нежно-розовый (словарь): 1 упоминаний
- светло-голубой (словарь): 1 упоминаний
- серо-буро-малиновый (авто): 1 упоминаний
- темно-коричневый (словарь): 1 упоминаний
- черной-черный (авто): 1 упоминаний
- ярко-желтый (словарь): 1 упоминаний

ЧАСТОТНОСТЬ ЦВЕТОВ:

ЦВЕТ	АБСОЛ	ОТН(слов)	ОТН(цвет)
светло-голубой	1	0.0455	0.1000
темно-коричневый	1	0.0455	0.1000
красный	1	0.0455	0.1000
нежно-розовый	1	0.0455	0.1000
ярко-желтый	1	0.0455	0.1000
серо-буро-малиновый	1	0.0455	0.1000 (авто)
красно-серо-оранжевый	1	0.0455	0.1000 (авто)
беловато-синевато-желтоватый	1	0.0455	0.1000 (авто)
космически-черный	1	0.0455	0.1000 (авто)
черной-черный	1	0.0455	0.1000 (авто)

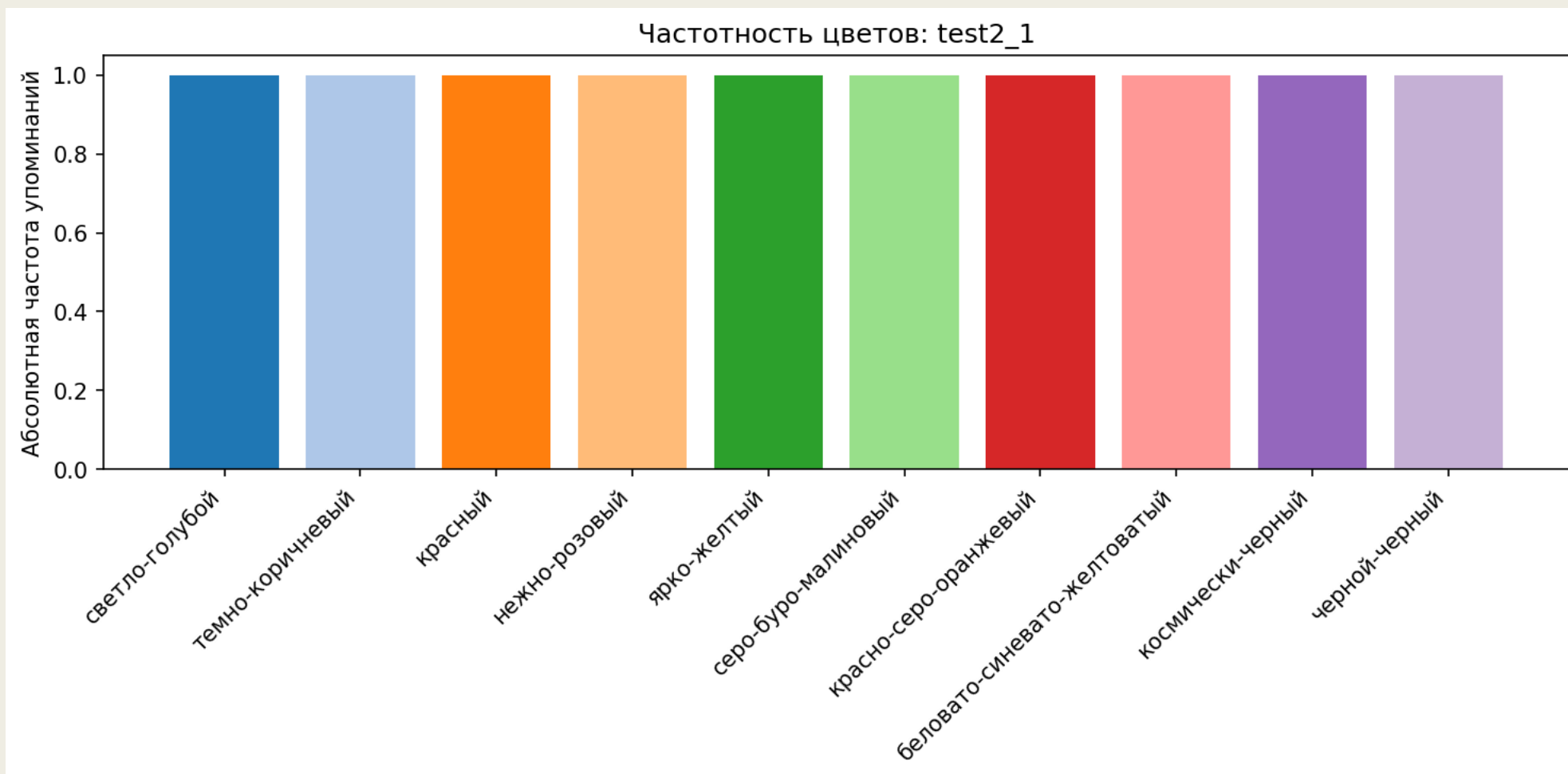
ВСЕГО ЦВЕТОВ: 10

Аня Ф.

Аня П.

Екатерина

График для теста 1 (Екатерина)



Тест 2(И.А. Бунин "Темные аллеи")

РЕЗУЛЬТАТЫ АНАЛИЗА

~~~~~  
Файл: bunin\_temnye\_all.txt

### ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Всего слов: 9411

Всего упоминаний цветов: 148

Плотность цветов относительно всех слов в тексте: 0.0157

### СТАТИСТИКА ПО ТИПАМ ЦВЕТОВ:

Уникальных цветов всего: 52

Наличие сложных цветов: Да

Простых: 46

Сложных: 6

Соотношение уникальных: простых 88.5% / сложных 11.5%

### ВСЕ НАЙДЕННЫЕ ЦВЕТА С ЧАСТОТАМИ:

(всего 52)

| Цвет                 | Абсолютная частота | Относительная частота(отн. всех слов) |
|----------------------|--------------------|---------------------------------------|
| алый                 | 2 раз(а)           | 0.000213                              |
| белоснежный          | 2 раз(а)           | 0.000213                              |
| белый                | 7 раз(а)           | 0.000744                              |
| бледный              | 3 раз(а)           | 0.000319                              |
| блестящий            | 2 раз(а)           | 0.000213                              |
| восковой             | 1 раз(а)           | 0.000106                              |
| гранатовый           | 1 раз(а)           | 0.000106                              |
| грязно-серый         | 1 раз(а)           | 0.000106                              |
| древесный            | 1 раз(а)           | 0.000106                              |
| желтенький           | 1 раз(а)           | 0.000106                              |
| желтый               | 2 раз(а)           | 0.000213                              |
| заколдованно-светлый | 1 раз(а)           | 0.000106                              |
| зеленоватый          | 1 раз(а)           | 0.000106                              |
| зеленый              | 5 раз(а)           | 0.000531                              |
| зеркальный           | 1 раз(а)           | 0.000106                              |
| золотистый           | 2 раз(а)           | 0.000213                              |
| золотой              | 4 раз(а)           | 0.000425                              |

|                    |           |          |
|--------------------|-----------|----------|
| канареечный        | 1 раз(а)  | 0.000106 |
| коралловый         | 1 раз(а)  | 0.000106 |
| коричневый         | 1 раз(а)  | 0.000106 |
| красный            | 8 раз(а)  | 0.000850 |
| кровавый           | 1 раз(а)  | 0.000106 |
| лиловый            | 1 раз(а)  | 0.000106 |
| лунный             | 2 раз(а)  | 0.000213 |
| медный             | 1 раз(а)  | 0.000106 |
| мутный             | 1 раз(а)  | 0.000106 |
| небесный           | 3 раз(а)  | 0.000319 |
| огненный           | 1 раз(а)  | 0.000106 |
| пестрый            | 3 раз(а)  | 0.000319 |
| прозрачный         | 2 раз(а)  | 0.000213 |
| радужный           | 1 раз(а)  | 0.000106 |
| ржавый             | 3 раз(а)  | 0.000319 |
| рубиновый          | 1 раз(а)  | 0.000106 |
| свежий             | 1 раз(а)  | 0.000106 |
| светлый            | 3 раз(а)  | 0.000319 |
| седой              | 2 раз(а)  | 0.000213 |
| серебряный         | 7 раз(а)  | 0.000744 |
| серенький          | 1 раз(а)  | 0.000106 |
| сере-серебристый   | 1 раз(а)  | 0.000106 |
| серый              | 11 раз(а) | 0.001169 |
| синий              | 2 раз(а)  | 0.000213 |
| смуглый            | 1 раз(а)  | 0.000106 |
| темный             | 18 раз(а) | 0.001913 |
| траурный           | 6 раз(а)  | 0.000638 |
| угольный           | 1 раз(а)  | 0.000106 |
| хрустально-золотой | 1 раз(а)  | 0.000106 |
| чернильный         | 1 раз(а)  | 0.000106 |
| черно-багровый     | 1 раз(а)  | 0.000106 |
| черно-смолистый    | 1 раз(а)  | 0.000106 |
| черный             | 18 раз(а) | 0.001913 |
| яркий              | 2 раз(а)  | 0.000213 |
| ясный              | 2 раз(а)  | 0.000213 |

# Тест 2(И.А. Бунин "Темные аллеи")

```
=====
АНАЛИЗ ТЕКСТА: Пользовательский файл: tests/bunin_temnye_all.txt
=====

ОБЩАЯ СТАТИСТИКА ТЕКСТА:
Общее количество слов: 9311
Количество уникальных словоформ: 4093

Общее количество слов-цветов: 136
Из них:
- простых цветов: 131
- сложных цветов: 5
Число разнообразных цветов: 47

СТАТИСТИКА ПО ЦВЕТАМ:
-----
```

|             |                  |                      |
|-------------|------------------|----------------------|
| тёмный      | Абс. частота: 18 | Отн. частота: 0.0019 |
| чёрный      | Абс. частота: 18 | Отн. частота: 0.0019 |
| серый       | Абс. частота: 11 | Отн. частота: 0.0012 |
| красный     | Абс. частота: 8  | Отн. частота: 0.0009 |
| белый       | Абс. частота: 7  | Отн. частота: 0.0008 |
| серебряный  | Абс. частота: 7  | Отн. частота: 0.0008 |
| траурный    | Абс. частота: 6  | Отн. частота: 0.0006 |
| золотой     | Абс. частота: 4  | Отн. частота: 0.0004 |
| бледный     | Абс. частота: 3  | Отн. частота: 0.0003 |
| небесный    | Абс. частота: 3  | Отн. частота: 0.0003 |
| ржавый      | Абс. частота: 3  | Отн. частота: 0.0003 |
| светлый     | Абс. частота: 3  | Отн. частота: 0.0003 |
| золотистый  | Абс. частота: 2  | Отн. частота: 0.0002 |
| седой       | Абс. частота: 2  | Отн. частота: 0.0002 |
| алый        | Абс. частота: 2  | Отн. частота: 0.0002 |
| прозрачный  | Абс. частота: 2  | Отн. частота: 0.0002 |
| блестящий   | Абс. частота: 2  | Отн. частота: 0.0002 |
| белоснежный | Абс. частота: 2  | Отн. частота: 0.0002 |

|                      |                 |                      |
|----------------------|-----------------|----------------------|
| яркий                | Абс. частота: 2 | Отн. частота: 0.0002 |
| лунный               | Абс. частота: 2 | Отн. частота: 0.0002 |
| ясный                | Абс. частота: 2 | Отн. частота: 0.0002 |
| синий                | Абс. частота: 2 | Отн. частота: 0.0002 |
| мутный               | Абс. частота: 1 | Отн. частота: 0.0001 |
| канареечный          | Абс. частота: 1 | Отн. частота: 0.0001 |
| огненный             | Абс. частота: 1 | Отн. частота: 0.0001 |
| древесный            | Абс. частота: 1 | Отн. частота: 0.0001 |
| восковой             | Абс. частота: 1 | Отн. частота: 0.0001 |
| кровавый             | Абс. частота: 1 | Отн. частота: 0.0001 |
| коричневый           | Абс. частота: 1 | Отн. частота: 0.0001 |
| рубиновый            | Абс. частота: 1 | Отн. частота: 0.0001 |
| чернильный           | Абс. частота: 1 | Отн. частота: 0.0001 |
| смуглый              | Абс. частота: 1 | Отн. частота: 0.0001 |
| коралловый           | Абс. частота: 1 | Отн. частота: 0.0001 |
| серенький            | Абс. частота: 1 | Отн. частота: 0.0001 |
| зеленоватый          | Абс. частота: 1 | Отн. частота: 0.0001 |
| гранатовый           | Абс. частота: 1 | Отн. частота: 0.0001 |
| грязно-серый         | Абс. частота: 1 | Отн. частота: 0.0001 |
| серо-серебристый     | Абс. частота: 1 | Отн. частота: 0.0001 |
| свежий               | Абс. частота: 1 | Отн. частота: 0.0001 |
| лиловый              | Абс. частота: 1 | Отн. частота: 0.0001 |
| хрустально-золотой   | Абс. частота: 1 | Отн. частота: 0.0001 |
| заколдованно-светлый | Абс. частота: 1 | Отн. частота: 0.0001 |
| черно-багровый       | Абс. частота: 1 | Отн. частота: 0.0001 |
| медный               | Абс. частота: 1 | Отн. частота: 0.0001 |
| зеркальный           | Абс. частота: 1 | Отн. частота: 0.0001 |
| радужный             | Абс. частота: 1 | Отн. частота: 0.0001 |
| угольный             | Абс. частота: 1 | Отн. частота: 0.0001 |

Самые часто встречающиеся цвета:

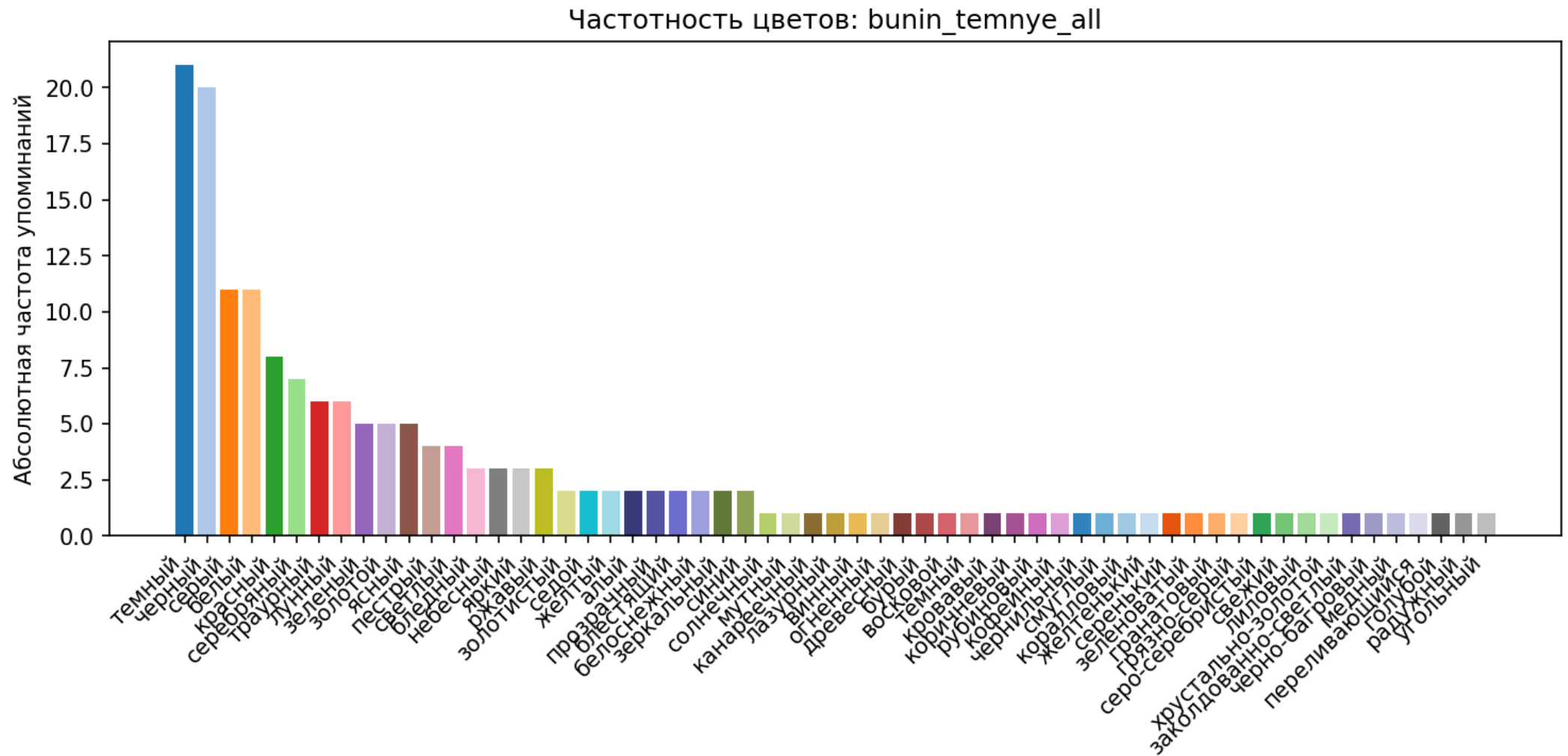
- тёмный (число упоминаний: 18)
- чёрный (число упоминаний: 18)

# Тест 2(И.А. Бунин "Темные аллеи")

|                                             |                     |       |           |           |                      |     |        |               |
|---------------------------------------------|---------------------|-------|-----------|-----------|----------------------|-----|--------|---------------|
| =====                                       | ЧАСТОТНОСТЬ ЦВЕТОВ: |       |           |           | огненный             | 1   | 0.0001 | 0.0057        |
| АНАЛИЗ ТЕКСТА: tests/bunin_temnye_all.txt   | ЦВЕТ                | АБСОЛ | ОТН(слов) | ОТН(цвет) | древесный            | 1   | 0.0001 | 0.0057        |
| =====                                       | -----               | ----- | -----     | -----     | бурый                | 1   | 0.0001 | 0.0057        |
| ОБЩАЯ СТАТИСТИКА ТЕКСТА:                    | темный              | 21    | 0.0022    | 0.1193    | восковой             | 1   | 0.0001 | 0.0057        |
| Всего слов: 9424                            | черный              | 20    | 0.0021    | 0.1136    | тёмный               | 1   | 0.0001 | 0.0057        |
| Уникальных слов(лемм): 2780                 | серый               | 11    | 0.0012    | 0.0625    | кровавый             | 1   | 0.0001 | 0.0057        |
| Общее упоминание цветов: 176                | белый               | 11    | 0.0012    | 0.0625    | коричневый           | 1   | 0.0001 | 0.0057        |
| Уникальных цветов: 59                       | красный             | 8     | 0.0008    | 0.0455    | рубиновый            | 1   | 0.0001 | 0.0057        |
| Плотность цветов: 0.0187                    | серебряный          | 7     | 0.0007    | 0.0398    | кофейный             | 1   | 0.0001 | 0.0057        |
| СТАТИСТИКА ПО ЦВЕТАМ:                       | траурный            | 6     | 0.0006    | 0.0341    | чернильный           | 1   | 0.0001 | 0.0057        |
| Самый частый цвет: 'темный' (21 упоминаний) | лунный              | 6     | 0.0006    | 0.0341    | смуглый              | 1   | 0.0001 | 0.0057        |
| СЛОЖНЫЕ ЦВЕТА (5):                          | зеленый             | 5     | 0.0005    | 0.0284    | коралловый           | 1   | 0.0001 | 0.0057        |
| - грязно-серый (словарь): 1 упоминаний      | золотой             | 5     | 0.0005    | 0.0284    | желтенький           | 1   | 0.0001 | 0.0057        |
| - заколдованно-светлый (авто): 1 упоминаний | ясный               | 5     | 0.0005    | 0.0284    | серенький            | 1   | 0.0001 | 0.0057        |
| - серо-серебристый (авто): 1 упоминаний     | пестрый             | 4     | 0.0004    | 0.0227    | зеленоватый          | 1   | 0.0001 | 0.0057        |
| - хрустально-золотой (авто): 1 упоминаний   | светлый             | 4     | 0.0004    | 0.0227    | гранатовый           | 1   | 0.0001 | 0.0057        |
| - черно-багровый (авто): 1 упоминаний       | бледный             | 3     | 0.0003    | 0.0170    | грязно-серый         | 1   | 0.0001 | 0.0057        |
|                                             | небесный            | 3     | 0.0003    | 0.0170    | серо-серебристый     | 1   | 0.0001 | 0.0057 (авто) |
|                                             | яркий               | 3     | 0.0003    | 0.0170    | свежий               | 1   | 0.0001 | 0.0057        |
|                                             | ржавый              | 3     | 0.0003    | 0.0170    | лиловый              | 1   | 0.0001 | 0.0057        |
|                                             | золотистый          | 2     | 0.0002    | 0.0114    | хрустально-золотой   | 1   | 0.0001 | 0.0057 (авто) |
|                                             | седой               | 2     | 0.0002    | 0.0114    | заколдованно-светлый | 1   | 0.0001 | 0.0057 (авто) |
|                                             | желтый              | 2     | 0.0002    | 0.0114    | черно-багровый       | 1   | 0.0001 | 0.0057 (авто) |
|                                             | алый                | 2     | 0.0002    | 0.0114    | медный               | 1   | 0.0001 | 0.0057        |
|                                             | прозрачный          | 2     | 0.0002    | 0.0114    | переливающийся       | 1   | 0.0001 | 0.0057        |
|                                             | блестящий           | 2     | 0.0002    | 0.0114    | голубой              | 1   | 0.0001 | 0.0057        |
|                                             | белоснежный         | 2     | 0.0002    | 0.0114    | радужный             | 1   | 0.0001 | 0.0057        |
|                                             | зеркальный          | 2     | 0.0002    | 0.0114    | угольный             | 1   | 0.0001 | 0.0057        |
|                                             | синий               | 2     | 0.0002    | 0.0114    | -----                |     |        |               |
|                                             | солнечный           | 1     | 0.0001    | 0.0057    | ВСЕГО ЦВЕТОВ:        | 176 |        |               |
|                                             | мутный              | 1     | 0.0001    | 0.0057    |                      |     |        |               |
|                                             | канареечный         | 1     | 0.0001    | 0.0057    |                      |     |        |               |
|                                             | лазурный            | 1     | 0.0001    | 0.0057    |                      |     |        |               |
|                                             | винный              | 1     | 0.0001    | 0.0057    |                      |     |        |               |
|                                             | огненный            | 1     | 0.0001    | 0.0057    |                      |     |        |               |

Екатерина

# График для теста 2(Екатерина)



# Недостатки и ограничения

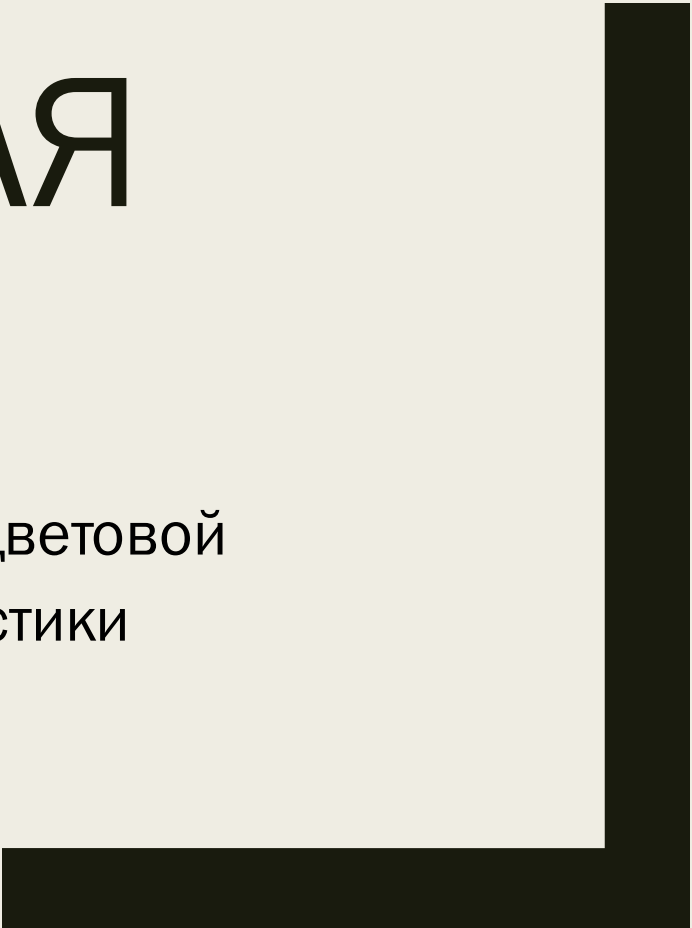
- Контекстуальная омонимия: например, слово «соломенный» распознается одинаково в разных контекстах: «соломенная шляпа» – название материала, «соломенные волосы» - метафорический цвет.(в "Темных аллеях" Бунина "ржавой крышей", "ржавые волосы", "в ржавом окладе")
- Метафорические цветообозначения: не распознаются конструкции типа цвета морской волны, оттенка спелой вишни (в "Темных аллеях" Бунина "глаза цвета желудя")
- Зависимость от качества лемматизации и морфологического анализа: Ошибки или особенности анализаторов могут привести к пропуску цвета (например rymorphu2 нормализует слова с приведением е к ё, из-за чего могут возникать проблемы или natasha распознала "голубей" как "голубой")



# ПРИКЛАДНАЯ ЗАДАЧА

Сравнительный анализ особенностей цветовой  
палитры в текстах различной стилистики

Анна Ф.



# Сравнение цветовой палитры произведений разных жанров И. А. Бунина.

- Критерии сравнения:
  - Общая насыщенность цветовой лексикой (% от всего текста)
  - Разнообразие палитры (количество уникальных цветов)
  - Сложность цветowych конструкций (% сложных цветов)
  - Доминирующие цветовые группы (ТОП-N самых частых)
- Объекты: Любовная проза «Тёмные аллеи», Лирическая проза «Антоновские яблоки», Философская притча «Господин из Сан-Франциско»
- Метрики: плотность цветовой лексики, разнообразие палитры, соотношение простых/сложных цветов, доминирующие цветовые группы.



# Результаты

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЦВЕТОВОЙ ПАЛИТРЫ ТЕКСТОВ

| Текст             | Темные аллеи | Антоновские яблоки | Господин из Сан-Франциско |
|-------------------|--------------|--------------------|---------------------------|
| Всего слов        | 9411         | 4239               | 6524                      |
| Упоминаний цветов | 148          | 93                 | 111                       |
| Плотность цветов  | 1.57%        | 2.19%              | 1.70%                     |
| Уникальные цвета  | 52           | 46                 | 57                        |
| Простые цвета     | 46           | 41                 | 48                        |
| Сложные цвета     | 6            | 5                  | 9                         |
| % простых         | 88.5%        | 89.1%              | 84.2%                     |
| % сложных         | 11.5%        | 10.9%              | 15.8%                     |



| ТОП-20 ЦВЕТОВ КАЖДОГО ТЕКСТА:                              |  |
|------------------------------------------------------------|--|
| Темные аллеи:                                              |  |
| 1. 'темный': 18 раз (0.1913%)                              |  |
| 2. 'черный': 18 раз (0.1913%)                              |  |
| 3. 'серый': 11 раз (0.1169%)                               |  |
| 4. 'красный': 8 раз (0.0850%)                              |  |
| 5. 'белый': 7 раз (0.0744%)                                |  |
| 6. 'серебряный': 7 раз (0.0744%)                           |  |
| 7. 'траурный': 6 раз (0.0638%)                             |  |
| 8. 'зеленый': 5 раз (0.0531%)                              |  |
| 9. 'золотой': 4 раз (0.0425%)                              |  |
| 10. 'бледный': 3 раз (0.0319%)                             |  |
| 11. 'небесный': 3 раз (0.0319%)                            |  |
| 12. 'ржавый': 3 раз (0.0319%)                              |  |
| 13. 'пестрый': 3 раз (0.0319%)                             |  |
| 14. 'светлый': 3 раз (0.0319%)                             |  |
| 15. 'золотистый': 2 раз (0.0213%)                          |  |
| 16. 'седой': 2 раз (0.0213%)                               |  |
| 17. 'желтый': 2 раз (0.0213%)                              |  |
| 18. 'алый': 2 раз (0.0213%)                                |  |
| 19. 'прозрачный': 2 раз (0.0213%)                          |  |
| 20. 'блестящий': 2 раз (0.0213%)                           |  |
| Частые простые: темный, черный, серый                      |  |
| Частые сложные: грязно-серый, серо-серебристый, хрустальны |  |

осподин из Сан-Франциско:

|                                                            |  |
|------------------------------------------------------------|--|
| 1. 'черный': 11 раз (0.1686%)                              |  |
| 2. 'красный': 8 раз (0.1226%)                              |  |
| 3. 'темный': 6 раз (0.0920%)                               |  |
| 4. 'золотой': 5 раз (0.0766%)                              |  |
| 5. 'серебряный': 5 раз (0.0766%)                           |  |
| 6. 'белый': 4 раз (0.0613%)                                |  |
| 7. 'бледный': 4 раз (0.0613%)                              |  |
| 8. 'синий': 4 раз (0.0613%)                                |  |
| 9. 'розовый': 3 раз (0.0460%)                              |  |
| 10. 'сизый': 3 раз (0.0460%)                               |  |
| 11. 'светлый': 3 раз (0.0460%)                             |  |
| 12. 'мраморный': 2 раз (0.0307%)                           |  |
| 13. 'огненный': 2 раз (0.0307%)                            |  |
| 14. 'мрачный': 2 раз (0.0307%)                             |  |
| 15. 'блестящий': 2 раз (0.0307%)                           |  |
| 16. 'смуглый': 2 раз (0.0307%)                             |  |
| 17. 'серый': 2 раз (0.0307%)                               |  |
| 18. 'оранжевый': 2 раз (0.0307%)                           |  |
| 19. 'белоснежный': 2 раз (0.0307%)                         |  |
| 20. 'свежий': 2 раз (0.0307%)                              |  |
| Частые простые: черный, красный, темный                    |  |
| Частые сложные: серо-зеленый, угольно-черный, светло-серый |  |

| Антоновские яблоки:                                     |  |
|---------------------------------------------------------|--|
| 1. 'черный': 11 раз (0.2595%)                           |  |
| 2. 'золотой': 6 раз (0.1415%)                           |  |
| 3. 'красный': 5 раз (0.1180%)                           |  |
| 4. 'ясный': 5 раз (0.1180%)                             |  |
| 5. 'свежий': 4 раз (0.0944%)                            |  |
| 6. 'солнечный': 4 раз (0.0944%)                         |  |
| 7. 'прохладный': 3 раз (0.0708%)                        |  |
| 8. 'белый': 3 раз (0.0708%)                             |  |
| 9. 'бирюзовый': 3 раз (0.0708%)                         |  |
| 10. 'прозрачный': 3 раз (0.0708%)                       |  |
| 11. 'желтый': 3 раз (0.0708%)                           |  |
| 12. 'бледный': 3 раз (0.0708%)                          |  |
| 13. 'кирпичный': 2 раз (0.0472%)                        |  |
| 14. 'багровый': 2 раз (0.0472%)                         |  |
| 15. 'темный': 2 раз (0.0472%)                           |  |
| 16. 'соломенный': 2 раз (0.0472%)                       |  |
| 17. 'голубой': 2 раз (0.0472%)                          |  |
| 18. 'мрачный': 2 раз (0.0472%)                          |  |
| 19. 'коралловый': 1 раз (0.0236%)                       |  |
| 20. 'земляной': 1 раз (0.0236%)                         |  |
| Частые простые: черный, золотой, красный                |  |
| Частые сложные: темно-синий, черно-лиловый, белый-белый |  |

# Выводы

- «Антоновские яблоки» демонстрирует высокую плотность теплых природных тонов, характерную для описательной лирической прозы.
- «Тёмные аллеи» используют сконцентрированную мрачную палитру, отражающую драматизм любовных отношений.
- «Господин из Сан-Франциско» отличается наибольшим разнообразием сложных символических оттенков, соответствующих философской глубине притчи.
- Плотность цветовой лексики максимальна в лирической прозе, где цвет выполняет в большей степени описательную функцию, и снижается в жанрах, где цвет несет символическую или эмоциональную нагрузку.
- Сложность цветовой палитры возрастает от лирической прозы к философской, что коррелирует с усложнением идейного содержания текстов.
- Эмоциональная тональность палитры различается: от теплой и светлой в лирической прозе к мрачной и контрастной в любовной и философской.

# ПРИКЛАДНАЯ ЗАДАЧА

«Отслеживание изменения настроения в  
художественных текстах по мере повествования»

Екатерина

# Тесты

- Ф.М. Достоевский «Преступление и наказание»
- Л.Н. Толстой «Война и мир»

# Выдержка цветовой дорожки по токенам

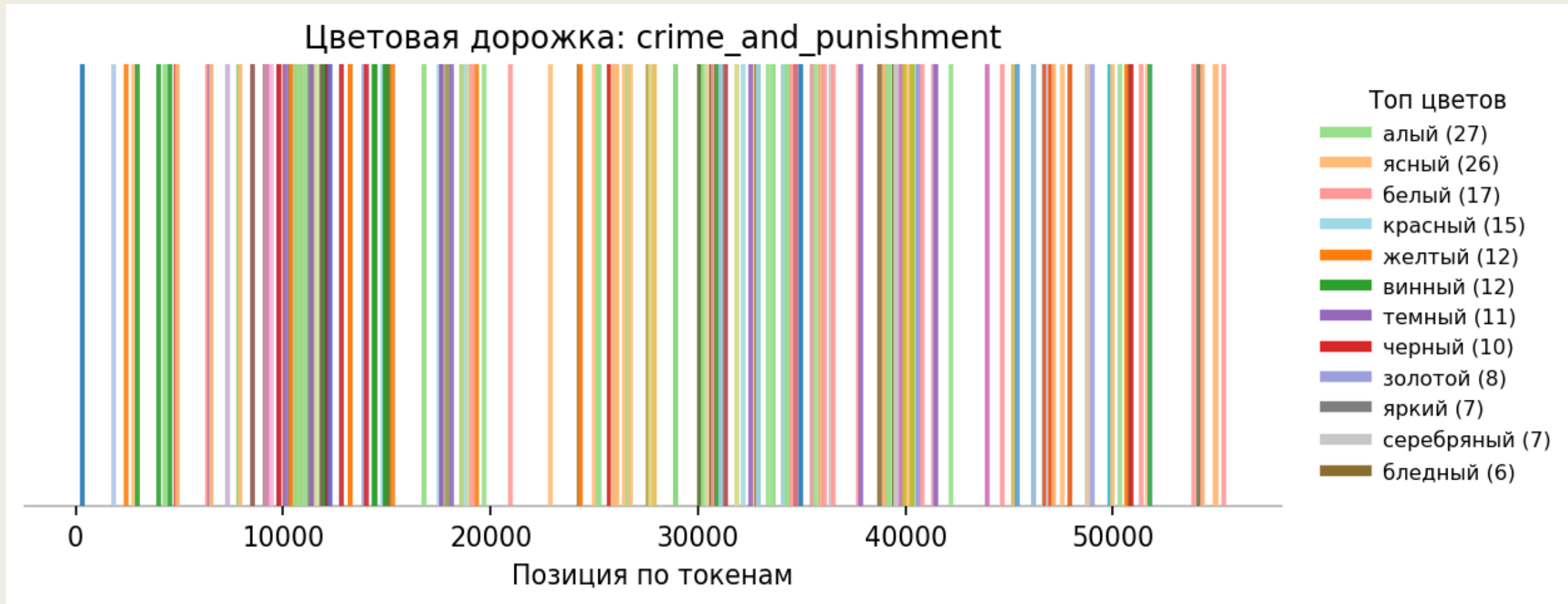
Преступление и наказание

|   | Позиция(токен) (1) | Цвет (2) |
|---|--------------------|----------|
| 1 | Позиция (токен)    | Цвет     |
| 2 | 295                | свежий   |
| 3 | 1835               | кровавый |
| 4 | 2402               | желтый   |
| 5 | 2768               | ясный    |
| 6 | 2944               | винный   |
| 7 | 3997               | винный   |
| 8 | 4322               | алый     |
| 9 | 4330               | алый     |

Война и мир

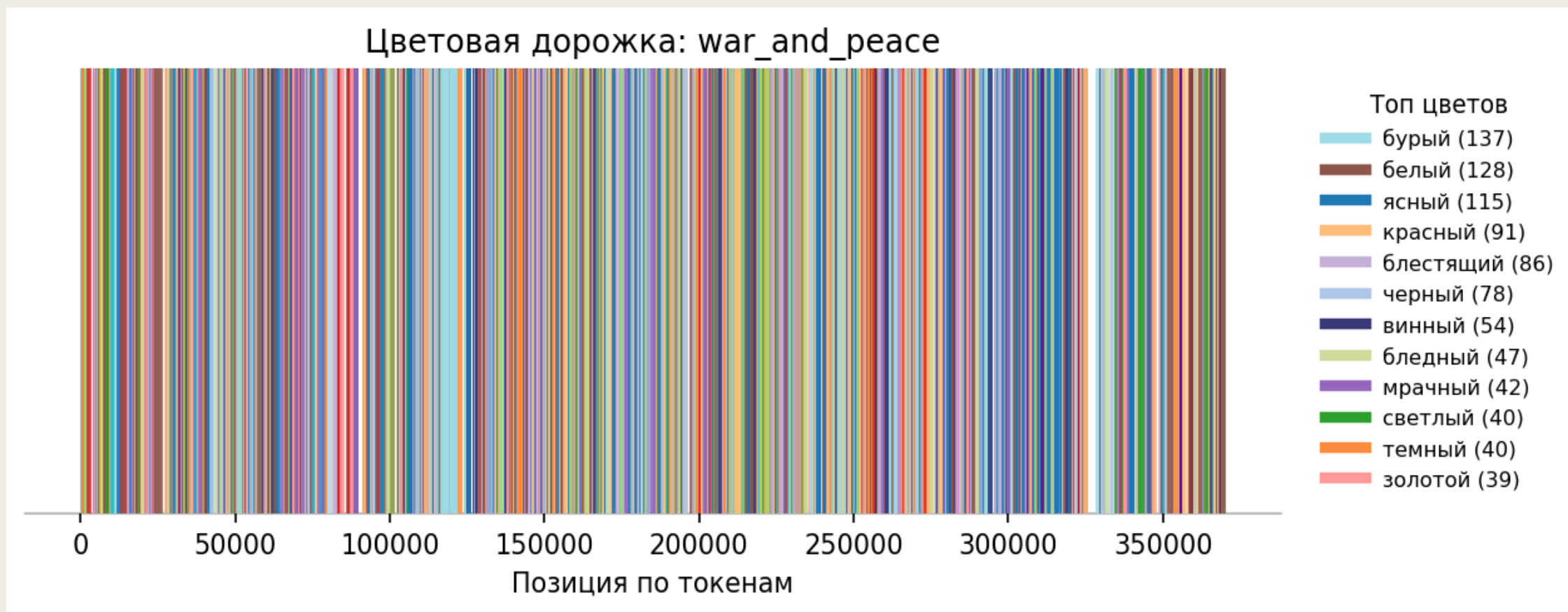
|   | Позиция(токен) (1) | Цвет (2) |
|---|--------------------|----------|
| 1 | Позиция (токен)    | Цвет     |
| 2 | 458                | ясный    |
| 3 | 564                | ясный    |
| 4 | 637                | ясный    |
| 5 | 752                | черный   |
| 6 | 873                | алый     |
| 7 | 1765               | красный  |
| 8 | 1846               | светлый  |
| 9 | 1919               | седой    |

# Цветовая дорожка текста для «Преступление и наказание»



Цвета часто противоположны, сменяются резко, как эмоциональные всплески. Преобладают красные и тёмные оттенки, что усиливает ощущение трагизма.

# Цветовая дорожка текста для «Война и мир»



В отличие от Достоевского, у Толстого цвета выстроены волнообразно: яркие эпизоды сменяются нейтральными, и композиция завершается светом.

# ВЫВОДЫ

- У Достоевского цвет служит усилителем эмоций: красные и чёрные оттенки концентрируются вокруг кульминации и передают внутреннюю борьбу.
- У Толстого цвет отражает течение жизни и настроение мира: частое появление светлых тонов создаёт ощущение равновесия и тепла.
- Оба автора используют цветовую символику как инструмент эмоционального построения текста, но делают это по-разному — через резкие контрасты у Достоевского и постепенные переходы у Толстого.



# ПРИКЛАДНАЯ ЗАДАЧА

«Анализ цветовых особенностей текстов разной тематики и построение хроматического профиля»

Анна П.

# Тесты

Текстовые файлы были разделены на 4 категории

1. Простые тесты (1-5)
2. Тексты по тематике "Природа" авторов-пейзажистов (6-9)
3. Научная фантастика (10-13)
4. Приключенческие романы о море (14-16)

# Тесты

- 1. Тест1
- 2. Тест2
- 3. Тест3
- 4. Тест4
- 5. Тест5
- 6. Иван Ефремов - Озеро горных духов
- 7. Михаил Пришвин - Кладовая солнца
- 8. Иван Бунин - Антоновские яблоки
- 9. Сергей Аксаков - Записки охотника
- 10. Иван Ефремов - Сердце змеи
- 11. Стругацкие - Попытка к бегству
- 12. Стругацкие - Далекая радуга
- 13. Станислав Лем - Солярис
- 14. Иван Ефремов - Катти Сарк
- 15. Жюль Верн - Дети капитана Гранта
- 16. Роберт Луис Стивенсон - Остров сокровищ

# По каждому тексту можно получить следующую статистику

```
=====
АНАЛИЗ ТЕКСТА: Иван Ефремов - Озеро горных духов
=====

ОБЩАЯ СТАТИСТИКА ТЕКСТА:
Общее количество слов: 4683
Количество уникальных словоформ: 2522

Общее количество слов-цветов: 91
Из них:
- простых цветов: 72
- сложных цветов: 19
Число разнообразных цветов: 45

СТАТИСТИКА ПО ЦВЕТАМ:
-----
```

|                  |                 |                      |
|------------------|-----------------|----------------------|
| голубой          | Абс. частота: 7 | Отн. частота: 0.0015 |
| прозрачный       | Абс. частота: 6 | Отн. частота: 0.0013 |
| сине-зелёный     | Абс. частота: 6 | Отн. частота: 0.0013 |
| красный          | Абс. частота: 5 | Отн. частота: 0.0011 |
| солнечный        | Абс. частота: 5 | Отн. частота: 0.0011 |
| яркий            | Абс. частота: 4 | Отн. частота: 0.0009 |
| белый            | Абс. частота: 3 | Отн. частота: 0.0006 |
| розовый          | Абс. частота: 3 | Отн. частота: 0.0006 |
| светящийся       | Абс. частота: 3 | Отн. частота: 0.0006 |
| кроваво-красный  | Абс. частота: 3 | Отн. частота: 0.0006 |
| серый            | Абс. частота: 2 | Отн. частота: 0.0004 |
| тусклый          | Абс. частота: 2 | Отн. частота: 0.0004 |
| мрачный          | Абс. частота: 2 | Отн. частота: 0.0004 |
| свежий           | Абс. частота: 2 | Отн. частота: 0.0004 |
| синий            | Абс. частота: 2 | Отн. частота: 0.0004 |
| зеленоватый      | Абс. частота: 2 | Отн. частота: 0.0004 |
| палевый          | Абс. частота: 2 | Отн. частота: 0.0004 |
| зеленовато-белый | Абс. частота: 2 | Отн. частота: 0.0004 |
| кровавый         | Абс. частота: 2 | Отн. частота: 0.0004 |
| ясный            | Абс. частота: 2 | Отн. частота: 0.0004 |
| пепельный        | Абс. частота: 2 | Отн. частота: 0.0004 |
| морской          | Абс. частота: 1 | Отн. частота: 0.0002 |
| свинцовый        | Абс. частота: 1 | Отн. частота: 0.0002 |
| бурый            | Абс. частота: 1 | Отн. частота: 0.0002 |

|                     |                 |                      |
|---------------------|-----------------|----------------------|
| темно-зелёный       | Абс. частота: 1 | Отн. частота: 0.0002 |
| чёрный              | Абс. частота: 1 | Отн. частота: 0.0002 |
| розоватый           | Абс. частота: 1 | Отн. частота: 0.0002 |
| светлый             | Абс. частота: 1 | Отн. частота: 0.0002 |
| пурпурный           | Абс. частота: 1 | Отн. частота: 0.0002 |
| изумрудный          | Абс. частота: 1 | Отн. частота: 0.0002 |
| загорелый           | Абс. частота: 1 | Отн. частота: 0.0002 |
| светло-жёлтый       | Абс. частота: 1 | Отн. частота: 0.0002 |
| желтоватый          | Абс. частота: 1 | Отн. частота: 0.0002 |
| синеvато-серый      | Абс. частота: 1 | Отн. частота: 0.0002 |
| серо-голубой        | Абс. частота: 1 | Отн. частота: 0.0002 |
| белоснежный         | Абс. частота: 1 | Отн. частота: 0.0002 |
| фиолетовый          | Абс. частота: 1 | Отн. частота: 0.0002 |
| желто-синий         | Абс. частота: 1 | Отн. частота: 0.0002 |
| тёмный              | Абс. частота: 1 | Отн. частота: 0.0002 |
| весенний            | Абс. частота: 1 | Отн. частота: 0.0002 |
| голубоватый         | Абс. частота: 1 | Отн. частота: 0.0002 |
| коричневато-красный | Абс. частота: 1 | Отн. частота: 0.0002 |
| сине-зеленоватый    | Абс. частота: 1 | Отн. частота: 0.0002 |
| голубовато-серый    | Абс. частота: 1 | Отн. частота: 0.0002 |
| мутный              | Абс. частота: 1 | Отн. частота: 0.0002 |

Самые часто встречающиеся цвета:  
- голубой (число упоминаний: 7)

```
=====
```

# Неточности, возникшие при анализе текстов

- В примере 10 самое частой слово - карий, но это не так, на самом деле Кари Рам - персонаж книги, а программа по ошибке считала его за цвет.

# Результаты (природа)

-----  
Тема: Природа – слова, встречающиеся в 4 из 4 текстов:

- белый
- голубой
- красный
- мрачный
- розовый
- свежий
- светлый
- синий
- солнечный
- тёмный
- чёрный
- ясный

-----  
Тема: Природа – слова, встречающиеся в 3 из 4 текстов:

- бурый
- весенний
- голубоватый
- загорелый
- золотистый
- золотой
- пепельный
- прозрачный
- рыжий
- свинцовый
- серый
- темно-зелёный

-----  
Тема: Природа – слова, встречающиеся в 2 из 4 текстов:

- белоснежный
- бледный
- блестящий
- болотный
- вишнёвый
- древесный
- желтоватый
- зеленоватый
- земляной
- кроваво-красный
- кровавый
- лиловый
- малиновый
- морской
- мужской
- мутный
- огненный
- палевый
- песчаный
- прохладный
- рыжеватый
- светло-жёлтый
- серебряный
- серенький
- сизый
- соломенный
- темно-красный
- тусклый
- яркий

# Результаты (научная фантастика)

-----  
Тема: Космос — слова, встречающиеся в 4 из 4 текстов:

- белый
- бледный
- блестящий
- голубой
- зеленоватый
- золотой
- красный
- матовый
- металлический
- огненный
- прозрачный
- серый
- тёмный
- чёрный
- яркий

-----  
Тема: Космос — слова, встречающиеся в 3 из 4 текстов:

- багровый
- белесый
- голубоватый
- зеркальный
- кровавый
- морской
- мрачный
- мутный
- оранжевый
- свежий
- светлый
- светящийся
- серебристый
- синий
- солнечный
- тусклый
- фиолетовый
- электронный
- ядовитый
- ясный

-----  
Тема: Космос — слова, встречающиеся в 2 из 4 текстов:

- бело-голубой
- бронзовый
- дымчатый
- желтоватый
- золотистый
- иссиня-чёрный
- красноватый
- кроваво-красный
- лиловый
- лунный
- монотонный
- мужской
- небесный
- полупрозрачный
- прохладный
- радужный
- ржавый
- розоватый
- розовый
- рыжий
- синеватый
- сиреневый
- спектральный
- стальной
- темно-зелёный
- темно-красный
- темно-синий
- угольно-чёрный
- фарфоровый

# Результаты (приключения про море)

-----  
Тема: Море — слова, встречающиеся в 3 из 3 текстов:

- белый
- красный
- медный
- морской
- свежий
- светлый
- свинцовый
- серебристый
- синий
- тёмный
- чёрный
- яркий
- ясный

-----  
Тема: Море — слова, встречающиеся в 2 из 3 текстов:

- багровый
- бледный
- блестящий
- болотный
- голубой
- древесный
- желтоватый
- загорелый
- земляной
- золотой
- кровавый
- лунный
- мрачный
- мутный
- небесный
- песчаный
- прозрачный
- прохладный
- рыжий
- сероватый
- серый
- солнечный
- соломенный
- тусклый
- угольный
- ядовитый



# Выводы

- Научная фантастика и текста про природу обладают в среднем большим количеством сложных цветов, чем приключенческая литература
- В литературе с описаниями природы больше природных оттенков (оттенки коричневого, жёлтого, зелёного)
- В научной фантастике больше тёмных холодных оттенков и ярких футуристических цветов (фиолетовый, ядовитый, электронный, зеркальный, серебристый)
- Почти в каждом приключенческом романе про море самое часто-употребляемое слово-цвет -- морской
- Существенный недостаток: существует много слов-цветов, относящихся к нескольким из этих категорий