

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова**

Факультет вычислительной математики и кибернетики

Кафедра алгоритмических языков

Отчёт по большому заданию

«Анализ неструктурированных данных»

Вариант 7. "Окрашенность"

**Студент: Китова Екатерина Денисовна
Группа: 324**

**Москва
2025 год**

Содержание

1 Постановка задачи	2
2 Реализация	2
2.1 Общая структура программы	3
2.2 Понятия простых и сложных цветов в программе	3
2.3 Функциональные блоки программы	4
3 Набор тестов и примеры работы	6
3.1 Тестовые данные	6
3.2 Таблицы частотного анализа	6
3.3 Графическая визуализация	7
3.4 Таблица сравнения для «Преступление и наказание» и «Война и мир»	10
3.5 Результаты работы сравнения «Преступление и наказание» и «Война и мир»	10
4 Содержательный анализ результатов	15
4.1 Основная часть	15
4.2 Дополнительная часть(прикладная задача)	16
5 Заключение и выводы	18

1 Постановка задачи

Необходимо было 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.

Для этого необходимо выполнить морфологический анализ словоформ текста, а также вычислить несколько общестатистических характеристик текста (общее число словоупотреблений, число различных, т.е. уникальных лемм и т.п.). Словарь различных цветов можно взять из Интернета, например: <https://textometr.ru/colors> или же составить его самостоятельно. Предлагается рассмотреть 2-5 художественных текстов разных авторов и сравнить полученные результаты.

Отчет: указание исследованных текстов, описание примененного словаря оценочных слов, программа с комментариями, подсчитанная статистика, пояснения к способу ее подсчета, выводы по сравнению.

Прикладные задачи: выявление особенностей стиля/ жанра/автора текста.

2 Реализация

Требовалось написать анализатор цветовой палитры текста. Данная программа выполняет лингвистический анализ русскоязычных текстов для оценки использования слов-названий цветов.

Также была сформулирована прикладная задача: "Отслеживание изменения настроения в художественных текстах по мере повествования"

Программа

- Находит в тексте упоминания цветов: простых(красный, желтый и т.п.), так и сложных(светло-голубой, лимонно-желтый и т.п.)
- Строит статистику частоты употребления каждого цвета(относительно всех слов в тексте(плотность) и всех найденных цветов)
- Определяет наиболее часто встречающиеся цвета
- Вычисляет общие лингвистические характеристики текста(общее кол-во слов, уникальных слов, упоминание цветов, уникальность и плотность цветов)
- Строит собственные словари цветов для каждого цвета
- Поддерживает автоматическое распознавание сложных цветов, которых нет в словаре

- Строит графики частот по тексту и сравнительные графики по нескольким текстам
- Формирует сравнительную таблицу по некоторым текстам (через ‘pandas’) и сохраняет её в ‘CSV’
- Строит таймлайн-график и таблицу(через ‘pandas’) появления цветов относительно повествования текста(прикладная задача)

Библиотеки

Для реализации программы использовались следующие библиотеки Python:

1. **natasha**: для морфологического анализа русского языка (лемматизация, определение частей речи)
2. **nltk**: для стемминга (приведения слов к основе)
3. **pandas**: для формирования сравнительных таблиц и экспорт в CSV
4. **matplotlib**: построение графиков
5. **re**: для работы с регулярными выражениями при предобработке текста
6. **collections**: для подсчета частотности слов
7. **pathlib**: работа с файлами и папками

2.1 Общая структура программы

Программа состоит из одного основного файла `color_analyzer.py`, где реализован класс `ColorAnalyzer`.

Каждая стадия анализа реализована отдельным методом — 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.

2.2 Понятия простых и сложных цветов в программе

Определим, что мы будем понимать под сложными и простыми цветами.

1. **Простые цвета** - это отдельные слова, обозначающие один цвет, без дефисов и дополнительных частей

Примеры: красный, синий, желтый, зеленый, белый, черный, розовый, бирюзовый, серый.

Простые цвета:

- (a) всегда присутствуют в словаре `dictionary.txt`;
- (b) могут иметь различные формы в тексте (например, желтый, желтые, желтым);

(c) программа приводит их к лемме (начальной форме) перед подсчётом.

2. **Сложные цвета** - это составные прилагательные, образованные через дефис из двух или более цветовых основ/простых лемм.

Они делятся на две группы:

(a) **Сложные из словаря**: составные цвета, внесённые в словарь ‘dictionary.txt’ в готовом виде: светло-голубой, темно-синий, лимонно-желтый, серебристо-серый и подобные. Если программа встречает их в тексте (в любой форме), они распознаются напрямую как известные цвета.

(b) **Автоматически определяемые сложные цвета**: 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.

Например: в словаре нет ‘нежно-лазурный’, но есть ‘лазурный’. Тогда программа будет искать в словаре простых цветов лемму от ‘лазурный’, тогда программа определит ‘нежно-лазурный’ как сложный цвет

2.3 Функциональные блоки программы

Разработанная программа реализует многоэтапный алгоритм анализа цветовой лексики, объединяющий методы морфологической обработки текста, лексикографического сопоставления и визуализации данных. Каждый функциональный блок выполняет свою часть анализа и тесно связан с другими.

1. **Загрузка и классификация словаря.** Первым этапом выполняется считывание файла `dictionary.txt`, содержащего список известных цветовых обозначений. Функция `_load_color_dictionary_from_file` автоматически разделяет все элементы на две категории:

- *простые цвета* (например, «синий», «жёлтый»), хранящиеся как отдельные леммы и их стемы (основы слов);
- *сложные цвета* (например, «бело-синий», «тёмно-зелёный»), которые анализируются как составные единицы.

Такое разделение позволяет эффективно сопоставлять найденные слова даже в изменённых формах и при наличии приставок или суффиксов.

2. **Предобработка текста.** Перед анализом текст нормализуется функцией `preprocess_text`: все символы приводятся к нижнему регистру, удаляются лишние пробелы и скрытые знаки, что обеспечивает корректную работу морфологического анализатора. 3. **Морфологический анализ.** На этом этапе используется библиотека `Natasha`, которая выполняет токенизацию и лемматизацию текста. Для каждого токена определяются:

- лемма (начальная форма слова);
- часть речи (POS-тег);
- основа слова (stem) — через модуль `nltk.SnowballStemmer`.

Это позволяет обрабатывать слова независимо от грамматической формы:

"красная-> "красный" зелёными-> "зелёный".

4. Распознавание сложных и составных цветов. Если токен содержит дефис, выполняется дополнительная логика:

1. если точное совпадение найдено в словаре — цвет учитывается напрямую;
2. если нет — анализируется последняя часть слова: если она совпадает с известным цветом, программа восстанавливает цветовую конструкцию;
3. в случае нескольких дефисов извлекается последняя лемма, к которой добавляется префикс, формируя итоговый «автоматически распознанный» цвет.

Например, слово «бело-сине-голубые» преобразуется в «бело-сине-голубой (авто)». Это решение позволяет учитывать цветовые сочетания, отсутствующие в словаре, но логически корректные с точки зрения русского языка.

5. Подсчёт статистики. Функция `calculate_statistics` собирает и агрегирует ключевые числовые показатели:

- общее количество слов и уникальных лемм;
- количество упоминаний цветов и их разнообразие;
- плотность цветовой лексики (отношение упоминаний цветов к общему числу слов);
- абсолютные и относительные частоты для каждого цвета;
- список наиболее частотных цветов.

Статистика выводится на экран и сохраняется в таблицы формата CSV.

6. Генерация и сохранение выходных данных. Для каждого анализируемого текста создаются:

- файл `colors_raw_<имя>.txt` — список всех уникальных цветов;
- файл `colors_freq_<имя>.csv` — таблица частотности (абсолютной и относительной);
- изображение `freq_<имя>.png` — бар-чарт распределения цветов.

Если анализируются несколько текстов, дополнительно формируется сравнительная таблица `comparison_table.csv` и сводный график `comparison_density.png`.

7. Визуализация данных. Программа реализует три основных типа графиков:

1. `plot_color_frequency` — показывает частоты употребления цветов в одном тексте;
2. `plot_comparison` — сопоставляет плотность цветовой лексики в разных произведениях;
3. `plot_color_timeline` — строит «цветовую дорожку» текста, отображающую последовательность появления цветовых слов.

В последнем случае каждому цвету соответствует вертикальная линия на временной оси токенов, а справа располагается легенда, автоматически сгенерированная на основе данных CSV.

8. Режимы анализа. Пользователь выбирает один из трёх режимов работы:

- `analyze_text` — анализ одного текста;
- `analyze_multiple_texts` — пакетная обработка нескольких произведений;
- `analyze_timeline` — построение и сохранение цветовой дорожки.

Каждый режим вызывает соответствующие функции и формирует собственные результаты в подпапке `figures/`.

3 Набор тестов и примеры работы

3.1 Тестовые данные

Для основного тестирования использовались два текста из классической русской литературы:

1. Ознакомительная часть романа Ф. М. Достоевского «Преступление и наказание»;
2. Первая версия романа Л. Н. Толстого «Война и мир»;

3.2 Таблицы частотного анализа

Таблица 1: Частоты цветовой лексики «Преступление и наказание»

Цвет	Абсолютная частота	Относит. к словам	Относит. к цветам
алый	27	0.000621	0.123853
ясный	26	0.000598	0.119266
белый	17	0.000391	0.077982
красный	15	0.000345	0.068807
желтый	12	0.000276	0.055046
вибранный	12	0.000276	0.055046
темный	11	0.000253	0.050459
черный	10	0.000230	0.045872
золотой	8	0.000184	0.036697

Таблица 2: Частоты цветовой лексики «Война и мир»

Цвет	Абсолютная частота	Относит. к словам	Относит. к цветам
бурый	137	0.000471	0.096547
белый	128	0.000440	0.090204

Цвет	Абсолютная частота	Относит. к словам	Относит. к цветам
ясный	115	0.000395	0.081043
красный	91	0.000313	0.064130
блестящий	86	0.000295	0.060606
черный	78	0.000268	0.054968
вишниный	54	0.000185	0.038055
бледный	47	0.000161	0.033122
мрачный	42	0.000144	0.029598
светлый	40	0.000137	0.028189

!В качестве примера показаны не все строки

3.3 Графическая визуализация

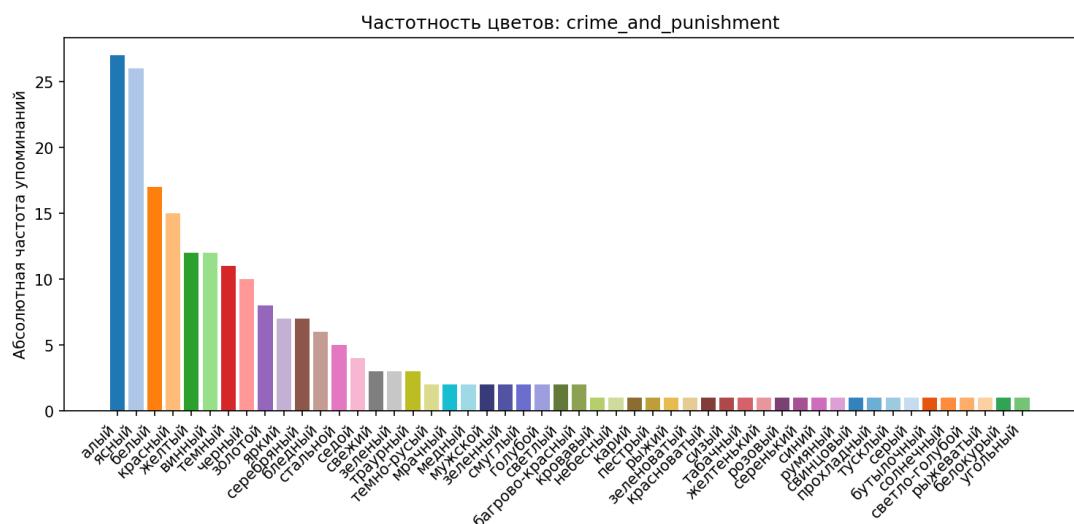


Рис. 1: График частотности цветов для «Преступление и наказание»

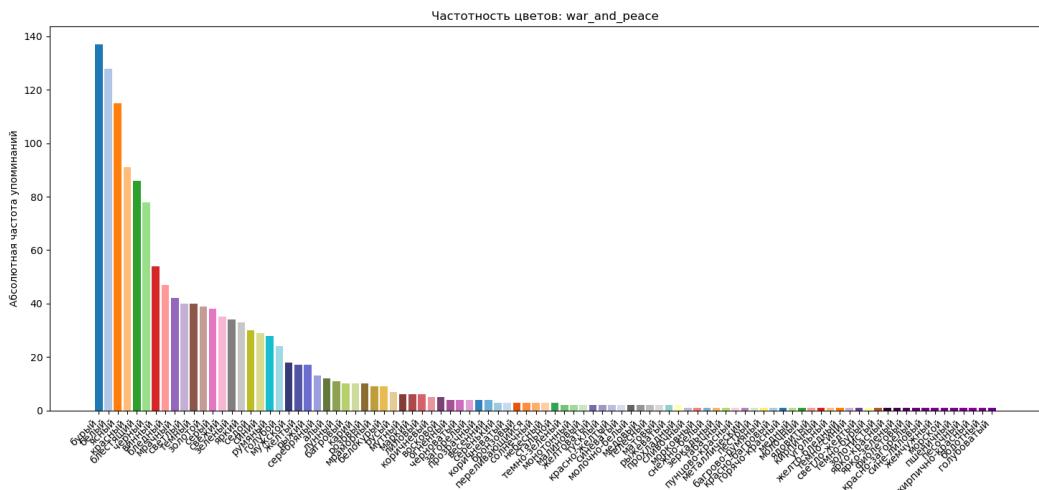


Рис. 2: График частотности цветов для «Война и мир»

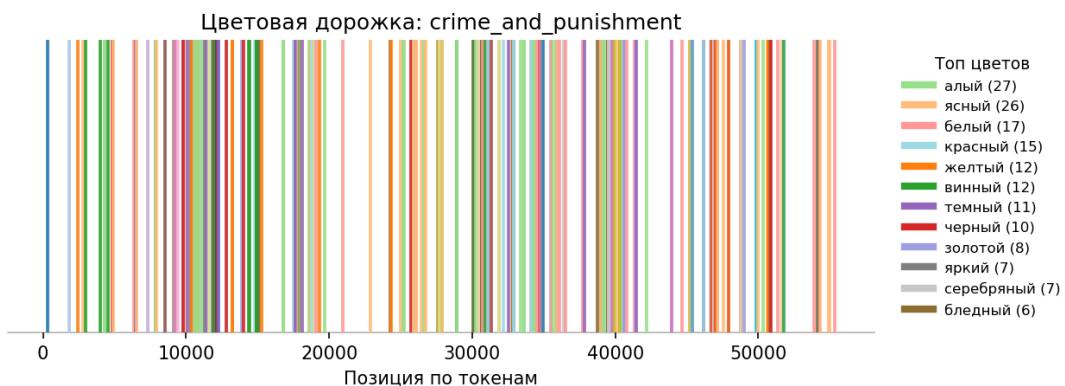


Рис. 3: Цветовая дорожка текста для «Преступление и наказание»

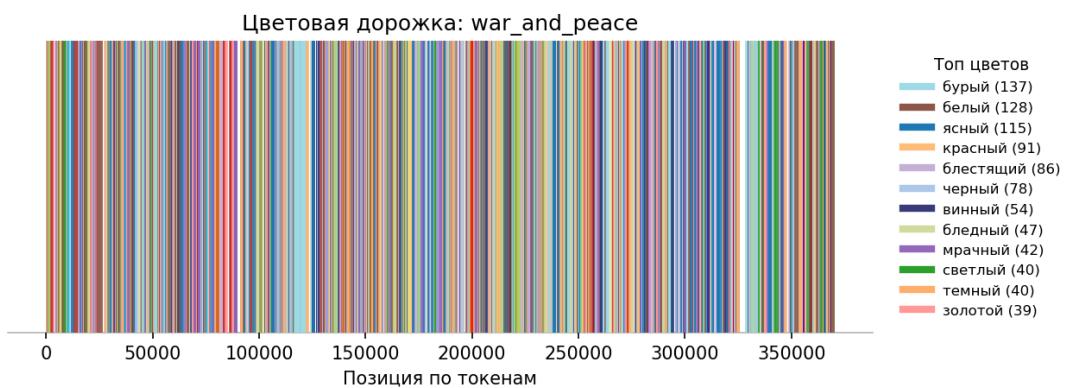


Рис. 4: Цветовая дорожка текста для «Война и мир»

Интерпретация: Каждый вертикальный штрих — позиция токена(ИМЕННО ТОКЕНА, НЕ СЛОВА) с обнаруженным цветом; Легенда упорядочена по частоте.

Таблица 3: Выдержка цветовой дорожки по токенам для «Преступление и наказание»

Позиция (токен)	Цвет
295	свежий
1835	кровавый
2402	жёлтый
2768	ясный
2944	винный
3997	винный
4322	алый
4330	алый
4538	винный
4852	чёрный
4902	ясный
6348	белый
6462	тёмный

Таблица 4: Выдержка цветовой дорожки по токенам для «Война и мир»

Позиция (токен)	Цвет
458	ясный
564	ясный
637	ясный
752	чёрный
873	алый
1765	красный
1846	светлый
1919	седой
2578	свежий
4629	золотой
4712	мрачный
4748	светлый
4751	блестящий

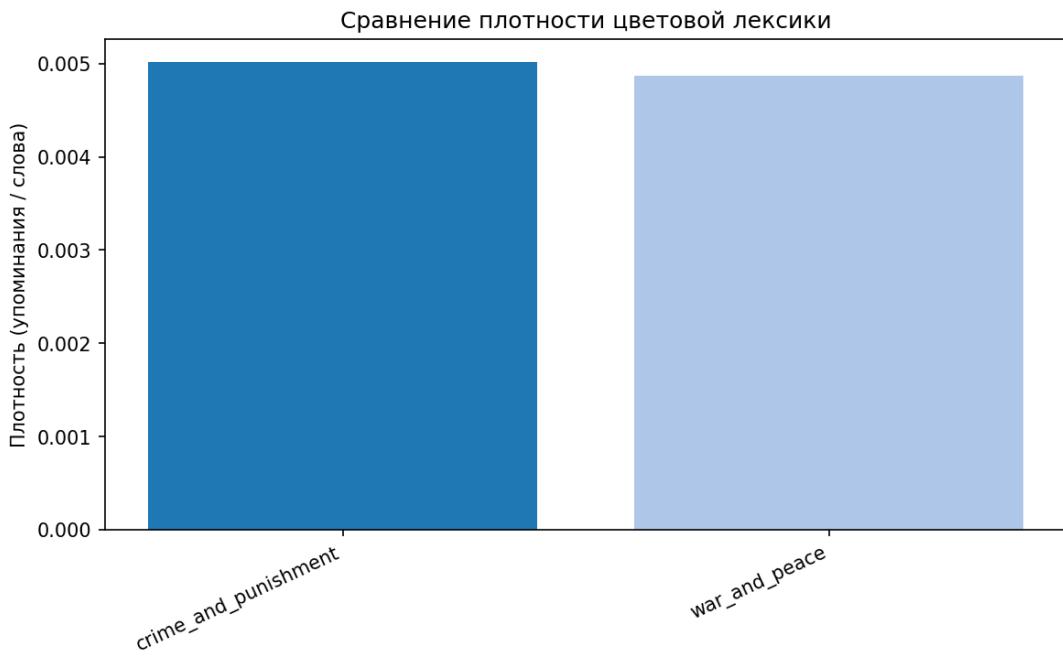


Рис. 5: Сравнение плотности цветовой лексики между текстами «Преступление и наказание» и «Война и мир»

3.4 Таблица сравнения для «Преступление и наказание» и «Война и мир»

Таблица 5: Сравнительная таблица по корпусам

Текст	Слова	Уник. слова	Упомин. цветов	Уник. цветов	Плотность	Топ-цвет(а)
tests/crime_and_punishment.txt	43495	6435	218	50	0.0050	алый
tests/war_and_peace.txt	291150	16505	1419	95	0.0049	бурый

3.5 Результаты работы сравнения «Преступление и наказание» и «Война и мир»

Выберите тип словаря:

- 1 - По умолчанию
- 2 - Свой словарь

Введите номер типа: 1

Загружено 482 цветов из файла dictionary.txt

Из них сложных цветов: 246

Простых цветов: 236

Выберите режим:

- 1 - Тестовый (один файл)
- 2 - Основной (несколько файлов)
- 3 - Цветовая дорожка (один файл)

Введите номер режима: 2

Введите пути к файлам через запятую

Файлы: tests/crime_and_punishment.txt, tests/war_and_peace.txt

=====

АНАЛИЗ ТЕКСТА: tests/crime_and_punishment.txt

=====

ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Всего слов: 43495

Уникальных слов(лемм): 6435

Общее упоминание цветов: 218

Уникальных цветов: 50

Плотность цветов: 0.0050

СТАТИСТИКА ПО ЦВЕТАМ:

Самый частый цвет: 'алый' (27 упоминаний)

СЛОЖНЫЕ ЦВЕТА (3):

- багрово-красный (авто): 2 упоминаний
- светло-голубой (словарь): 1 упоминаний
- темно-русый (словарь): 2 упоминаний

ЧАСТОТНОСТЬ ЦВЕТОВ:

ЦВЕТ	АБСОЛ	ОТН(слов)	ОТН(цвет)
алый	27	0.0006	0.1239
ясный	26	0.0006	0.1193
белый	17	0.0004	0.0780
красный	15	0.0003	0.0688
желтый	12	0.0003	0.0550
вишненый	12	0.0003	0.0550
черный	10	0.0002	0.0459
темный	10	0.0002	0.0459
золотой	8	0.0002	0.0367
яркий	7	0.0002	0.0321
серебряный	7	0.0002	0.0321
бледный	6	0.0001	0.0275
стальной	5	0.0001	0.0229
зеленый	5	0.0001	0.0229
седой	4	0.0001	0.0183
свежий	3	0.0001	0.0138
траурный	3	0.0001	0.0138
темно-русый	2	0.0000	0.0092
мрачный	2	0.0000	0.0092
медный	2	0.0000	0.0092
мужской	2	0.0000	0.0092
смуглый	2	0.0000	0.0092

голубой	2	0.0000	0.0092
светлый	2	0.0000	0.0092
багрово-красный	2	0.0000	0.0092 (авто)
кровавый	1	0.0000	0.0046
небесный	1	0.0000	0.0046
карий	1	0.0000	0.0046
пестрый	1	0.0000	0.0046
рыжий	1	0.0000	0.0046
зеленоватый	1	0.0000	0.0046
красноватый	1	0.0000	0.0046
сизый	1	0.0000	0.0046
тёмный	1	0.0000	0.0046
табачный	1	0.0000	0.0046
желтенький	1	0.0000	0.0046
розовый	1	0.0000	0.0046
серенький	1	0.0000	0.0046
синий	1	0.0000	0.0046
румяный	1	0.0000	0.0046
свинцовый	1	0.0000	0.0046
прохладный	1	0.0000	0.0046
тусклый	1	0.0000	0.0046
серый	1	0.0000	0.0046
бутылочный	1	0.0000	0.0046
солнечный	1	0.0000	0.0046
светло-голубой	1	0.0000	0.0046
рыжеватый	1	0.0000	0.0046
белокурый	1	0.0000	0.0046
угольный	1	0.0000	0.0046

ВСЕГО ЦВЕТОВ: 218

Список уникальных цветов сохранён: figures/colors_raw_crime_and_punishment.txt

Частоты цветов сохранены: figures/colors_freq_crime_and_punishment.csv

График сохранён: figures/freq_crime_and_punishment.png

=====

АНАЛИЗ ТЕКСТА: tests/war_and_peace.txt

=====

ОБЩАЯ СТАТИСТИКА ТЕКСТА:

Всего слов: 291150

Уникальных слов(лемм): 16505

Общее упоминание цветов: 1419

Уникальных цветов: 95

Плотность цветов: 0.0049

СТАТИСТИКА ПО ЦВЕТАМ:

Самый частый цвет: 'бурый' (137 упоминаний)

СЛОЖНЫЕ ЦВЕТА (17):

- багрово-румяный (авто): 1 упоминаний
- горячо-красный (авто): 1 упоминаний
- желто-бледный (авто): 1 упоминаний
- кирпично-красный (словарь): 1 упоминаний
- красно-багровый (авто): 1 упоминаний
- красно-желтый (словарь): 2 упоминаний
- красно-загорелый (авто): 1 упоминаний
- молочно-белый (словарь): 2 упоминаний
- пунцово-красный (авто): 1 упоминаний
- светло-желтый (словарь): 1 упоминаний
- сине-лиловый (авто): 1 упоминаний
- снежно-белый (авто): 1 упоминаний
- темно-бурый (словарь): 1 упоминаний
- темно-зеленый (словарь): 2 упоминаний
- темно-синий (словарь): 1 упоминаний
- ярко-зеленый (словарь): 1 упоминаний
- ярко-красный (словарь): 1 упоминаний

ЧАСТОТНОСТЬ ЦВЕТОВ:

ЦВЕТ	АБСОЛ	ОТН(слов)	ОТН(цвет)
бурый	137	0.0005	0.0965
белый	128	0.0004	0.0902
ясный	115	0.0004	0.0810
красный	91	0.0003	0.0641
блестящий	86	0.0003	0.0606
черный	78	0.0003	0.0550
вишневый	54	0.0002	0.0381
бледный	47	0.0002	0.0331
мрачный	42	0.0001	0.0296
светлый	40	0.0001	0.0282
темный	40	0.0001	0.0282
золотой	39	0.0001	0.0275
серый	38	0.0001	0.0268
свежий	35	0.0001	0.0247
зеленый	34	0.0001	0.0240
яркий	33	0.0001	0.0233
седой	30	0.0001	0.0211
синий	29	0.0001	0.0204
румяный	28	0.0001	0.0197
голубой	24	0.0001	0.0169
мужской	18	0.0001	0.0127
желтый	17	0.0001	0.0120
рыжий	17	0.0001	0.0120
серебряный	13	0.0000	0.0092
алый	12	0.0000	0.0085
лунный	11	0.0000	0.0078

багровый	10	0.0000	0.0070
карий	10	0.0000	0.0070
розовый	10	0.0000	0.0070
мраморный	9	0.0000	0.0063
белокурый	9	0.0000	0.0063
русый	7	0.0000	0.0049
мутный	6	0.0000	0.0042
маркий	6	0.0000	0.0042
лиловый	6	0.0000	0.0042
коричневый	5	0.0000	0.0035
восковой	5	0.0000	0.0035
огненный	4	0.0000	0.0028
черноватый	4	0.0000	0.0028
загорелый	4	0.0000	0.0028
прозрачный	4	0.0000	0.0028
весенний	4	0.0000	0.0028
серенький	3	0.0000	0.0021
коричневатый	3	0.0000	0.0021
бронзовый	3	0.0000	0.0021
переливающийся	3	0.0000	0.0021
солнечный	3	0.0000	0.0021
небесный	3	0.0000	0.0021
стальной	3	0.0000	0.0021
темно-зеленый	2	0.0000	0.0014
монотонный	2	0.0000	0.0014
малиновый	2	0.0000	0.0014
желтоватый	2	0.0000	0.0014
тусклый	2	0.0000	0.0014
красно-желтый	2	0.0000	0.0014
синеватый	2	0.0000	0.0014
молочно-белый	2	0.0000	0.0014
медовый	2	0.0000	0.0014
телесный	2	0.0000	0.0014
матовый	2	0.0000	0.0014
рыжеватый	2	0.0000	0.0014
прохладный	2	0.0000	0.0014
сливочный	1	0.0000	0.0007
морковный	1	0.0000	0.0007
снежно-белый	1	0.0000	0.0007 (авто)
зеркальный	1	0.0000	0.0007
табачный	1	0.0000	0.0007
пунцово-красный	1	0.0000	0.0007 (авто)
металлический	1	0.0000	0.0007
пестрый	1	0.0000	0.0007
багрово-румяный	1	0.0000	0.0007 (авто)
красно-багровый	1	0.0000	0.0007 (авто)
горячо-красный	1	0.0000	0.0007 (авто)
медный	1	0.0000	0.0007

молочный	1	0.0000	0.0007
ядовитый	1	0.0000	0.0007
кирпичный	1	0.0000	0.0007
угольный	1	0.0000	0.0007
желто-бледный	1	0.0000	0.0007 (авто)
темно-синий	1	0.0000	0.0007
светло-желтый	1	0.0000	0.0007
темно-бурый	1	0.0000	0.0007
золотистый	1	0.0000	0.0007
ярко-красный	1	0.0000	0.0007
ярко-зеленый	1	0.0000	0.0007
фиолетовый	1	0.0000	0.0007
красно-загорелый	1	0.0000	0.0007 (авто)
сине-лиловый	1	0.0000	0.0007 (авто)
жемчужный	1	0.0000	0.0007
морской	1	0.0000	0.0007
пшеничный	1	0.0000	0.0007
песочный	1	0.0000	0.0007
кирично-красный	1	0.0000	0.0007
болотный	1	0.0000	0.0007
голубоватый	1	0.0000	0.0007

ВСЕГО ЦВЕТОВ: 1419

Список уникальных цветов сохранён: figures/colors_raw_war_and_peace.txt

Частоты цветов сохранены: figures/colors_freq_war_and_peace.csv

График сохранён: figures/freq_war_and_peace.png

=====

СРАВНИТЕЛЬНАЯ ТАБЛИЦА:

=====

	Текст	Слова	Уник.	слова	Упомин.	цветов	Уник.	цветов	Пло-
tests/crime_and_punishment.txt	43495		6435		218			50	
tests/war_and_peace.txt	291150		16505		1419			95	

CSV со сравнительной таблицей сохранён: figures/comparison_table.csv

График сравнений сохранён: figures/comparison_density.png

4 Содержательный анализ результатов

Благодаря нашей программе, теперь можно провести содержательный анализ текстов.

4.1 Основная часть

Общие наблюдения Оба текста показывают близкую плотность цветовой лексики — примерно одно упоминание цвета на каждые 200 слов. Это значит, что у обоих авторов цветовые описания занимают сопоставимую долю текста. Однако различается то, как распределены цвета. У Достоевского 50 уникальных цветов на

218 упоминаний — то есть каждый цвет встречается в среднем 4–5 раз. У Толстого 95 уникальных при 1 419 упоминаниях — в среднем около 15 раз на цвет. Следовательно, Достоевский использует меньше оттенков, но более разнообразно, а Толстой — шире, но чаще повторяет одни и те же цвета.

Доминирующие цвета

- В «Преступлении и наказании» лидирует **алый**, за ним идут **ясный, белый, красный, жёлтый**. Цвета в основном контрастные и эмоционально насыщенные — красные, белые и тёмные. Это подчёркивает напряжённую атмосферу и внутренние конфликты.
- В «Войне и мире» преобладают **бурый, белый, ясный, красный, блестящий, чёрный**. Палитра мягче и «земнее»: рядом с контрастными красными часто встречаются естественные оттенки — бурый, серый, золотой. Это создаёт впечатление разнообразного, но уравновешенного мира.

Сложные цвета Программа выделила также составные (дефисные) цвета. У Достоевского найдено всего 3 таких оттенка (например, *багрово-красный, светло-голубой, темно-русый*), а у Толстого — аж 17 (в том числе *молочно-белый, ярко-красный, тёмно-бурый*). То есть Толстой чаще комбинирует цветовые описания, создавая более детальные образы, а у Достоевского сложные оттенки используются точечно — как акценты.

Сравнение

- У **Достоевского** цветовая лексика более эмоциональна и контрастна. Преобладают красно-белые сочетания с вкраплениями чёрного и тёмного. Это визуально и символически поддерживает темы крови, страдания, очищения.
- У **Толстого** палитра шире, но спокойнее. Часты естественные и светлые тона (*белый, ясный, золотой, бурый*), создающие ощущение жизненного цикла и гармонии. Красные и тёмные цвета присутствуют, но не доминируют.

Выводы

1. Оба автора активно используют цветовую лексику, но по-разному: у Достоевского — контраст и драма, у Толстого — равновесие и многослойность.
2. «Преступление и наказание» отличается большей эмоциональной плотностью, где каждый цвет несёт сильную смысловую нагрузку.
3. «Война и мир» демонстрирует разнообразие палитры, но с мягкими переходами и повторяющимися мотивами, отражающими масштаб и многообразие жизни.

4.2 Дополнительная часть(прикладная задача)

Наша прикладная задача была сформулирована так:

"Отслеживание изменения настроения в художественных текстах по мере повествования"

Проведем детальный анализ таймлайна для каждого произведения.

«Преступление и наказание»

- Уже в начале (*свежий*, 295) появляется спокойный оттенок, но вскоре он сменяется на *кровавый* (1835), что отражает нарастание тревоги и напряжения.
- В середине (2400–4300) цвета постоянно колеблются между светлыми и тяжёлыми: *жёлтый — ясный — винный — алый*. Это визуально передаёт внутреннюю борьбу Раскольникова — между просветлением и виной.
- К концу последовательности появляются *чёрный* и *тёмный*, а перед ними — *белый*. Такая смена (белый => тёмный) символизирует короткое облегчение, за которым вновь следует тяжесть.

В целом картина напряжённая и контрастная: цвета часто противоположны, сменяются резко, как эмоциональные всплески. Преобладают красные и тёмные оттенки, что усиливает ощущение трагизма.

«Война и мир»

- Начало (458–637) отмечено тройным *ясный* — символ спокойствия и гармонии. Далее следуют *чёрный* и *алый*, добавляющие контраст и динамику, но они не разрушают общий баланс.
- Средняя часть (1765–2578) чередует *красный*, *светлый*, *седой*, *свежий*. Здесь появляется разнообразие тонов — от тёплых к холодным, что отражает движение между действием и раздумьем.
- Завершение (4629–4751) — *золотой* => *мрачный* => *светлый* => *блестящий*. Это будто визуальный переход от кульминации к умиротворённому финалу.

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

Сравнение

- У Достоевского цвета сосредоточены ближе к середине — в моменты наибольшего внутреннего кризиса. Они насыщенные и конфликтные (алый, винный, чёрный), что создаёт ощущение взрыва эмоций.
- У Толстого цвета распределены равномернее по тексту. Светлые и тёплые оттенки (*ясный*, *светлый*, *золотой*) встречаются чаще и «смягчают» общую тональность.
- Таким образом, «*Преступление и наказание*» — это история цвета через контраст, а «*Война и мир*» — через плавное чередование и гармонию.

Выводы

1. У Достоевского цвет служит усилителем эмоций: красные и чёрные оттенки концентрируются вокруг кульминации и передают внутреннюю борьбу.
2. У Толстого цвет отражает течение жизни и настроение мира: частое появление светлых тонов создаёт ощущение равновесия и тепла.
3. 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.

5 Заключение и выводы

Проведённая работа позволила создать и применить программу для анализа цветовой лексики в художественных текстах, показавшую, что автоматизация подобных исследований может быть не только технически осуществима, но и содержательно плодотворна. Сочетание количественного подхода с литературоведческим анализом дало возможность по-новому взглянуть на использование цвета как смыслового элемента в произведениях классической литературы.

Разработанная программа успешно обрабатывает тексты любой длины, автоматически выделяет прилагательные, обозначающие цвета, учитывая сложные формы, строит таблицы частотности и визуализирует результаты в виде диаграмм и цветовых дорожек. Такой подход позволил не просто подсчитать количество упоминаний цветов, но и выявить их распределение по тексту, динамику и функциональное назначение. Использование трёх режимов работы — одиночного, сравнительного и динамического — сделало систему гибким инструментом для различных задач анализа.

Сравнение двух произведений, «Преступления и наказания» Ф. М. Достоевского и «Войны и мира» Л. Н. Толстого, показало, что при близкой плотности цветовых упоминаний (0.005) авторы используют цветовую лексику с разными целями. У Достоевского цвета служат отражением внутреннего конфликта, психологического напряжения и драматизма — доминируют контрастные оттенки (алый, чёрный, белый, винный). У Толстого же цветовая палитра шире и мягче: в ней преобладают природные и спокойные тона (бурый, золотой, светлый, ясный), создающие ощущение гармонии и жизненного равновесия. Эти различия отражают не только индивидуальный стиль писателей, но и их философское мировосприятие — у Достоевского внутренняя борьба человека с самим собой, у Толстого — единство человека и мира.

Построенные цветовые дорожки позволили увидеть, как динамика употребления цветовых слов соотносится с эмоциональной структурой текста. В «Преступлении и наказании» всплески насыщенных оттенков совпадают с кульминационными моментами, усиливая драматическое напряжение. В «Войне и мире» же цветовая динамика распределена равномерно, отражая многоплановость повествования и естественный ритм жизни героев.

Таким образом, проведённое исследование подтверждает, что цветовая лексика играет важную роль в построении художественного мира произведения. Её анализ позволяет количественно описывать эмоциональные и стилистические особенности текста, а также визуализировать их развитие во времени. В перспективе данный подход можно расширить на тексты разных эпох и языков, что позволит создать масштабную базу данных цветовой символики и проследить эволюцию художественного восприятия цвета в литературе.