

Анализ сочинений в формате ЕГЭ.  
Соотношение числа уникальных слов, средней длины  
предложения и оценки по критериям 6 и 9.  
*(на материале сочинений, размещенных на сайте  
Kritika.24)*

Целевая  
аудитория

- Учителя, выпускники,  
составители  
аналитических отчетов в  
системе образования

## Задачи

- 1. Составление корпуса текстов.
- 2. Обработка материала и выявление соотношений: K6 (точность и выразительность речи), количества уникальных слов, средней длины предложения; K9 (соблюдение грамматических норм) и средней длины предложения.
- 3. Бонус: составление облака слов (с учетом и без учета фраз-клише)

## Этапы работы

1. Парсинг сайта, составление сводного файла с текстами и **df**.
2. Препроцессинг
3. Лемматизация
4. Анализ корреляции K6 (точность и выразительность речи) и а) общего количества уникальных слов; б) средней длины предложения.
5. Анализ соотношения средней длины предложения и оценки по K9 (грамматические ошибки).



**КРИТИКА24.RU**

Верный  
помощник!

Поделиться с друзьями

РЕГИСТРАЦИЯ

ВХОД

Логин

Пароль

Не запоминать ▾

ОК

Вход через VK

[забыли пароль?](#)

Проверка сочинений

Заказать сочинение

МЕНЮ

Писатели по алфавиту

Краткие содержания

По произведениям

Биографии

Анализы стихов

Сравнительные

Страницы: 1 2 3 4 5 6 7 8 ... 296

### Другие сочинения

- 1) текст Бунина про слепого. Если выйти на мол, встретишь, несмотря на яркое солнце, резкий ветер; 2) текст Батыгина про любовь. Прокопия Ивановича привезли в больницу глубокой ночью. Его провожала жена (**Оценка эксперта: 38 из 48**)

Сочинение 1: Когда возможно равенство между людьми? Этому вопросу посвящен текст И.А.Бунина. Чтобы привлечь внимание читателей к поставленной проблеме, автор пишет, что только «смерть или великие...

2. [Для чего человеку нужен друг? По текстам Бондарева и Астафьева](#) (**Оценка эксперта: 37 из 50**)

1. Для чего человеку нужен друг? Данную проблему поднимает Бондарев в своём тексте. Рассуждая над проблемой, автор описывает дом, в котором живет старенький морщинистый человек. Тот дом всегда...

3. [Как детство влияет на жизнь человека?](#) (**Оценка эксперта: 33 из 38**)

Что означает слово детство? Как оно влияет на жизнь человека? Детство — это смех, радость, веселье, познание чего-нибудь нового. Юностью называется период, когда ты будто в «розовых очках». Одним...

4. [Как изменились отношения между родителями и детьми в последние годы?](#) (**Оценка эксперта: 25 из 26**)

Как изменились отношения между родителями и детьми в последние годы ? Ответ на этот вопрос можно найти в

# Корпус – 1524 сочинения

	Ссылка	Заголовок	Текст сочинения	Критерий К6	Критерий К9
0	<a href="https://www.kritika24.ru/page.php?id=41596">https://www.kritika24.ru/page.php?id=41596</a>	По тексту С.Мизерова «Каждый ли человек способе	2	2	2
1	<a href="https://www.kritika24.ru/page.php?id=108339">https://www.kritika24.ru/page.php?id=108339</a>	Стоит ли верить в добрс	Сочувствие к человеку и вер	2	2
2	<a href="https://www.kritika24.ru/page.php?id=139992">https://www.kritika24.ru/page.php?id=139992</a>	Проблема сохранения с	В предложенном тексте Ф. К	2	2
3	<a href="https://www.kritika24.ru/page.php?id=122199">https://www.kritika24.ru/page.php?id=122199</a>	Наконец я приехал в Од	Какую роль играет прощени	2	2
4	<a href="https://www.kritika24.ru/page.php?id=147852">https://www.kritika24.ru/page.php?id=147852</a>	Угасание жизни русской	В предложенном тексте А. И	2	2
5	<a href="https://www.kritika24.ru/page.php?id=140795">https://www.kritika24.ru/page.php?id=140795</a>	Проблема роли плагиат	В представленном для анали	2	2
6	<a href="https://www.kritika24.ru/page.php?id=135809">https://www.kritika24.ru/page.php?id=135809</a>	Как дети видят мир? По	Как дети видят мир? Именн	2	2
7	<a href="https://www.kritika24.ru/page.php?id=151864">https://www.kritika24.ru/page.php?id=151864</a>	Как изменяется отноше	Как изменяется отношение	2	2
8	<a href="https://www.kritika24.ru/page.php?id=152319">https://www.kritika24.ru/page.php?id=152319</a>	Можно ли говорить об	Можно ли говорить об оши	2	2
9	<a href="https://www.kritika24.ru/page.php?id=151829">https://www.kritika24.ru/page.php?id=151829</a>	Искренность — хорошо	Искренность — хорошо или	2	2
10	<a href="https://www.kritika24.ru/page.php?id=147768">https://www.kritika24.ru/page.php?id=147768</a>	Всегда ли достижение ч	Всегда ли достижение чело	2	2
11	<a href="https://www.kritika24.ru/page.php?id=146447">https://www.kritika24.ru/page.php?id=146447</a>	Проблема развития чув	В предложенном тексте И. Ф	2	2
12	<a href="https://www.kritika24.ru/page.php?id=152310">https://www.kritika24.ru/page.php?id=152310</a>	Отношение человека к	В предложенном для анали	2	2
13	<a href="https://www.kritika24.ru/page.php?id=148297">https://www.kritika24.ru/page.php?id=148297</a>	В чём смысл жизни? По	В чём смысл жизни? Этот во	2	1
14	<a href="https://www.kritika24.ru/page.php?id=104716">https://www.kritika24.ru/page.php?id=104716</a>	Как человек находит ду	Как человек находит духовн	2	2
15	<a href="https://www.kritika24.ru/page.php?id=109507">https://www.kritika24.ru/page.php?id=109507</a>	Как человек относится	Как человек относится к при	2	2
16	<a href="https://www.kritika24.ru/page.php?id=150490">https://www.kritika24.ru/page.php?id=150490</a>	Дружба по тексту Дж. Р	Почему дружба важна для ч	1	1
17	<a href="https://www.kritika24.ru/page.php?id=148160">https://www.kritika24.ru/page.php?id=148160</a>	Проблема веры в чело	Насколько важна вера в чел	2	2
18	<a href="https://www.kritika24.ru/page.php?id=146144">https://www.kritika24.ru/page.php?id=146144</a>	Сквернословие по текст	Как противостоять сквернос	2	2
19	<a href="https://www.kritika24.ru/page.php?id=148469">https://www.kritika24.ru/page.php?id=148469</a>	Проблема отношения ч	Человечество с момента свс	2	2
20	<a href="https://www.kritika24.ru/page.php?id=116384">https://www.kritika24.ru/page.php?id=116384</a>	Проблема нравственно	В предложенном для анали	2	2
21	<a href="https://www.kritika24.ru/page.php?id=147274">https://www.kritika24.ru/page.php?id=147274</a>	Роль перемен в жизни ч	В предложенном тексте Д. А	2	2
22	<a href="https://www.kritika24.ru/page.php?id=103021">https://www.kritika24.ru/page.php?id=103021</a>	Как человек должен дей	Как человек должен действ	2	2
23	<a href="https://www.kritika24.ru/page.php?id=116397">https://www.kritika24.ru/page.php?id=116397</a>	19 В избушке у самого л	В предложенном для анали	2	2
24	<a href="https://www.kritika24.ru/page.php?id=127464">https://www.kritika24.ru/page.php?id=127464</a>	По тексту Л. А. Евсеевой	В предложенном для анали	2	2
25	<a href="https://www.kritika24.ru/page.php?id=146725">https://www.kritika24.ru/page.php?id=146725</a>	Какие взгляды имеет че	Какие взгляды имеет члове	2	2
26	<a href="https://www.kritika24.ru/page.php?id=139499">https://www.kritika24.ru/page.php?id=139499</a>	Проблема роли добрых	В современном мире редко	1	2
27	<a href="https://www.kritika24.ru/page.php?id=151587">https://www.kritika24.ru/page.php?id=151587</a>	Проблема сострадания	Нужно ли помогать людям, п	2	2

# Парсинг

## Сложности и особенности

- На первом этапе собрать материал со всех страничек, не отбирать ссылки, при переходе по которым не будет сочинения + нужны были не все сочинения, а только с первых 26 страниц

```
list_ = []
for link in soup.find_all("a"):
    list_.append((str(link.get('href')), link.text))
for link, title in list_:
    index = link.find('php?id=')
    #print(link[index + 7:])
    try:
        number = int(link[index + 7:]) #по кол-ву символом (слово 'page' тоже учитывалось)
    except ValueError:
        number = -1
    if index != -1 and link.find('EGJe') == -1 and number > 4917: #задать промежуток: != от 0 до 4917
        ege_link_and_title.append(('https://www.kritika24.ru/' + link, title))
```

99. «На пути к океану» Ф. Кривин «Маленькая рыбка Анабас живет далеко от морей...»  
(Оценка эксперта: 24 из 24)

Может ли человек изменить свою жизнь? Ставя перед собой какую-либо цель, хватит ли у него силы преодолеть все препятствия, чтобы достичь её? Над этими вопросами заставляет нас задуматься автор,...

100. В чем заключается героизм на войне? Первое впечатление ваше непременно самое неприятное (Оценка эксперта: 24 из 24)

В чем заключается героизм на войне? Именно над этим вопросом задумывается русский писатель Л.Н.Толстой. Проследим, как автор раскрывает сюжетную линию. Л.Н.Толстой отмечает поведение солдата в...

Страницы: 1 2 3 4 5 6 7 8 ... 296

### Полезные статьи О заработке на Критика24.ру

- Как правильно написать название сочинения?
- Как получить максимальную прибыль и от чего она зависит?
- Откуда берутся посещения и как их увеличить?
- F.A.Q. ответы на частые вопросы

# Парсинг

## Сложности и особенности

- Часть сочинений пришлось исключить: вместо текста была картинка (удаляла по тегу); встретилась некорректная сумма баллов (первые три сочинения).

```
soup = BeautifulSoup(page.text, 'lxml')
flag_img = False #убираем все ссылки на изображения: тег img + src=screen'
for elem in soup.find_all('img'):
    if elem['src'].startswith('screen'):
        #print(elem['src'])
        flag_img = True
if flag_img:
    continue
```



# Парсинг

## Сложности и особенности

- Парсить иногда приходилось по несколько страниц, т.к. сайт отказывался работать (выставленная задержка не помогала).
- Затем, чтобы понимать, что вообще происходит был использован следующий способ:

```
for link, title in ege_link_and_title:
    if counter % (len_percent + 1) == len_percent:
        print(f'06работан{ending(int(100 * counter / len_))} {int(100 * counter / len_)}% сочинений\n')
        counter += 1
    flag_page = False # загружена ли страница (да/нет)
    while not flag_page:
        try:
            page = rq.get(link)
        except Exception:
            time.sleep(1)
            continue
        flag_page = True
```

# Парсинг

The screenshot displays a Python IDE interface. On the left, a file explorer shows a directory named 'Детские сочинения' containing various files including '.py', '.xlsx', '.txt', and '.png' files. The main editor window shows a Python function 'def ending(number):' with conditional logic. The console window at the bottom shows the execution progress of a script, displaying the percentage of processed essays.

```
def ending(number):  
    if number % 100 == 11:  
        return 'o'  
    number %= 10  
    if number == 1:  
        return ''  
    else:  
        return 'o'
```

url = 'https://www.kritika24.ru/page.php?id=1907&top=kritik&'

Run: ЕГЭ. Анализ x 2. ЕГЭ (проба) x

- Обработан 41% сочинений
- Обработано 42% сочинений
- Обработано 43% сочинений
- Обработано 44% сочинений
- Обработано 45% сочинений
- Обработано 46% сочинений

# Парсинг

## Сложности и особенности

- Вынесение K9 и K6 в отдельные столбцы

```
essay = '' #создали строку
flag_K = False #работаем с критериями: отбираем по тегу с текстом K6 и K9
k_6 = None
k_9 = None
for elem in soup.find_all('p'):
    if not flag_K:
        if elem.text.find('K1') == -1: #не найдено K1, еще идет текст
            essay += elem.text
        else:
            flag_K = True #если нашел K1, дальше идут критерии, в текст не пишет
    else:
        if elem.text.find('K6') != -1:
            score_list = re.findall(r'[0-2]', elem.text)
```

```
if len(score_list) > 0: #нам нужен 1-й элемент, т.к. запись мб разной: 2 из 2 или 2/2
    k_6 = score_list[0]

if elem.text.find('K9') != -1:
    score_list = re.findall(r'[0-2]', elem.text)
    if len(score_list) > 0:
        k_9 = score_list[0]
```

# Работа с ТЕКСТОВЫМ файлом и df

```
ege_essays = pd.read_csv("essays3.csv", encoding="utf-8", sep=',')
index_name = ege_essays.columns[0]
ege_essays = ege_essays.set_index(index_name)
#print(ege_essays)

list_df_k6 = [ege_essays[ege_essays['Критерий К6'] == i] for i in range(3)]
list_df_k9 = [ege_essays[ege_essays['Критерий К9'] == i] for i in range(3)]
```

## КРИТЕРИЙ 6

<b>Точность и выразительность речи</b>	
Работа экзаменуемого характеризуется точностью выражения мысли, разнообразием грамматического строя речи.	2
<b>* Высший балл по этому критерию экзаменуемый получает только в случае, если высший балл получен по критерию K10</b>	
Работа экзаменуемого характеризуется точностью выражения мысли, <b>но</b> прослеживается однообразие грамматического строя речи, <b>или</b> работа экзаменуемого характеризуется разнообразием грамматического строя речи, <b>но</b> есть нарушения точности выражения мысли	1
Работа экзаменуемого отличается бедностью словаря и однообразием грамматического строя речи	0

## Связь количества уникальных лексем и оценки по K6

```
for df_ in list_df_k6:
    sum_ = 0
    # for row in df_.itertuples():
    #     print(row['Текст сочинения'], end='\n')
    for number in range(len(df_)):
        new_text = lemmatized_texts(df_.iloc[number]['Текст сочинения'], stop_words)
        sum_ += unique_words_amount(new_text)
    #print(len(df_))
    average.append(sum_/len(df_))
for i in range(len(average)):
    print(f'Средняя количество уникальных слов в сочинениях с критерием K6 {i}: {average[i]}\n')
```

# Связь количества уникальных лексем и оценки по K6

Среднее количество уникальных слов в сочинениях с критерием K6 0: 131.20626631853787

Среднее количество уникальных слов в сочинениях с критерием K6 1: 131.75782881002087

Среднее количество уникальных слов в сочинениях с критерием K6 2: 138.87027027027028

## Связь количества уникальных лексем и оценки по K6

- В стоп-слова были включены слова из клише (фразы, которые рекомендуют использовать в тексте сочинения для организации речи )

Среднее количество уникальных слов в сочинениях с критерием K6 0: 95.73629242819844

Среднее количество уникальных слов в сочинениях с критерием K6 1: 96.08768267223383

Среднее количество уникальных слов в сочинениях с критерием K6 2: 102.5027027027027



## Соотношение средней длины предложения и количества баллов по К6

```
def average_sentence_len(text):  
    #print(text)  
    sentence_list = re.split('[а-я][\.\!?\']', text) #будет съедать последнюю букву в слове, но для нас это не так важно  
    sum_ = 0 #длина, сколько слов во всех предложениях текста  
    amount_ = 0 #сколько всего предложений в тексте  
    #print(sentence_list)  
    for sentence in sentence_list:  
        sum_ += len(sentence.split()) #разбили на слова и добавили в переменную длину предложения (кол-во слов)  
        if len(sentence.split()) > 0: #убрали пустые предложения без слов  
            amount_ += 1 #считаем предложения  
    #print(sum_, amount_)  
    return sum_ / amount_ #среднее
```

```
ege_essays = pd.read_csv("essays3.csv", encoding="utf-8", sep=',') #обращаемся к датафрейму  
index_name = ege_essays.columns[0]  
ege_essays = ege_essays.set_index(index_name)  
#print(ege_essays)  
  
list_df_k6 = [ege_essays[ege_essays['Критерий К6'] == i] for i in range(3)] #разбиваем по оценкам (3 датафрейма по балл  
list_df_k9 = [ege_essays[ege_essays['Критерий К9'] == i] for i in range(3)]
```

- 1) количество слов в предложении текста/на общее количество предложений
- 2) разбить по оценке от 0 до 2 баллов; ранжировать сочинения, посчитать среднее количество в каждой категории

# Соотношение средней длины предложения и количества баллов по K6

Среднее количество слов в предложении в сочинениях с критерием K6 0: 17.01559701420924

Среднее количество слов в предложении в сочинениях с критерием K6 1: 16.849384546666883

Среднее количество слов в предложении в сочинениях с критерием K6 2: 16.45811733162175

Process finished with exit code 0

# КРИТЕРИЙ 9

Соблюдение грамматических норм	
Грамматических ошибок нет	2
Допущены одна-две ошибки	1
Допущено три и более ошибок	0

# Соотношение средней длины предложения и количества баллов по K9

Среднее количество слов в предложении в сочинениях с критерием K9 0: 16.720126885472652

Среднее количество слов в предложении в сочинениях с критерием K9 1: 16.76083612516923

Среднее количество слов в предложении в сочинениях с критерием K9 2: 16.94158200573256

Process finished with exit code 0

## Промежуточные выводы

- К6: сочинения, получившие 2 балла отличаются большим количеством уникальных слов (слова-клише практически не влияют на данный показатель) и меньшей длиной предложения. Ученики с меньшим словарным запасом стараются использовать в предложении больше слов (вообще предложения очень длинные: в пробном тексте среднее количество доходило до 20-ти) .
- К9: практически нет зависимости от длины предложения. Скорее всего, ошибки допускаются не в самой структуре предложения, а в словоформах.

# Облако слов

- 2 облака: второе – без стоп-слов + слов из списка с клише (список предварительно лемматизирован)

```
stop_words = stopwords.words('russian')
stop_words.extend(['критерий', 'к6', 'к9', 'который', 'это', 'к6', 'к9', 'из-за', 'он', 'её', 'его', 'их', 'ее'])

lemmatized_texts_ = lemmatized_texts(texts, stop_words)
wordcloud = WordCloud().generate(' '.join(lemmatized_texts_))
plt.imshow(wordcloud) # Что изображаем
plt.axis("off") # Без подписей на осях
plt.show() # показать изображение
print(wordcloud)
```

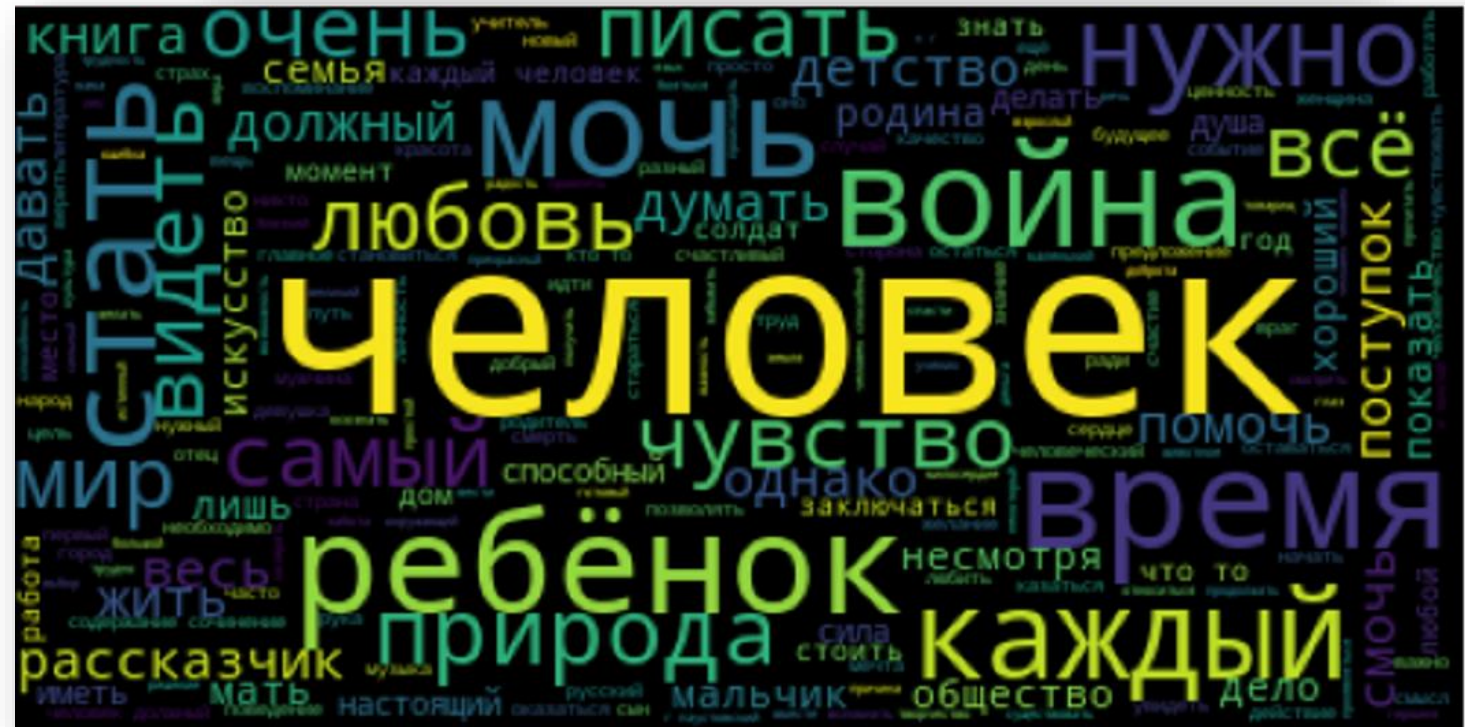
```
new_stop_words = stop_words.copy()
new_stop_words.extend(lemmatized_texts(cliches, stop_words))
```

```
new_lemmatized_texts = lemmatized_texts(texts, new_stop_words)
new_wordcloud = WordCloud().generate(' '.join(new_lemmatized_texts))
plt.imshow(new_wordcloud) # Что изображаем
plt.axis("off") # Без подписей на осях
plt.show() # показать изображение
print(new_wordcloud)
```



- Облако слов  
(все 1524  
сочинения)

- 
- Человек-друг
- жизнь мир герой любовь каждый война являться дополнять друг пример автор позиция мнение чувство свой ребенок нужно привести говорить её автор считать думать понимать жизнь душа иметь маленько проблема каждый война являться дополнять друг пример автор позиция мнение чувство свой ребенок нужно привести говорить её автор считать думать понимать



Возможные  
дальнейшие  
направления  
изучения

- Работа с реальными сочинениями школьников (на сайтах мы не можем отследить, кто автор, кто оценивает сочинение)
- Детальное составление списка клише: сравнение существующих в сборниках для подготовки с программами из текстов (для получения более объективной картины)
- Найти наиболее частотные слова и сочетания в открытых материалах ФИПИ, вариантах прошлых лет для установления статистической вероятности встречаемости той или иной темы (более актуально для ОГЭ)
- Сравнение с текстами других жанров.