

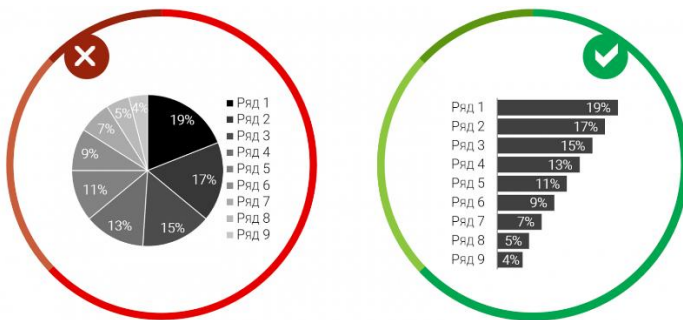
## Лабораторная работа №8

1)

**Задание (Уровни 1, 2, 3): законспектировать основные рекомендации по созданию визуализации**

### ПРАВИЛО 1. ПРАВИЛЬНЫЙ ТИП ГРАФИКА

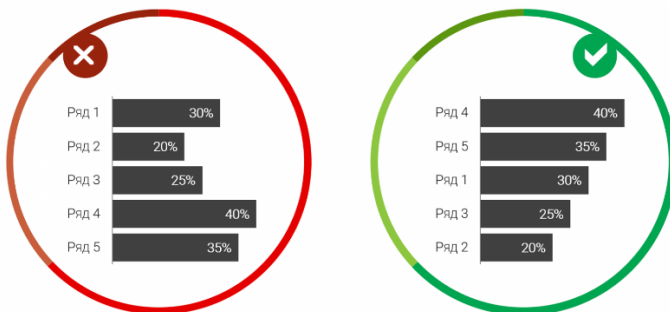
Используйте правильный тип и формат визуализации.



Главная цель визуализации — упростить и ускорить восприятие информации. Выбранный формат и тип графика должны этому способствовать, а не мешать.

### ПРАВИЛО 2. ЛОГИЧЕСКИЙ ПОРЯДОК

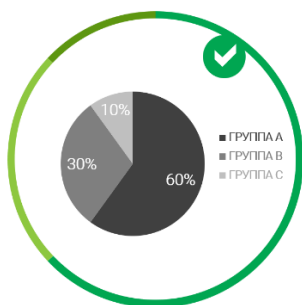
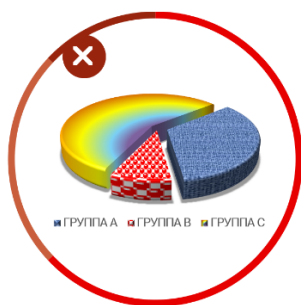
Располагайте данные логично.



Обязательно располагайте данные в логическом порядке. Чаще всего это последовательно от большего к меньшему.

### ПРАВИЛО 3. ПРОСТОЙ ДИЗАЙН

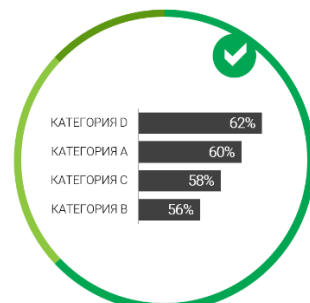
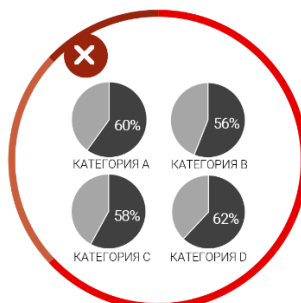
Дизайн не должен препятствовать пониманию или искажать данные.



Избегайте бесполезных элементов дизайна, таких как градиенты, тени, эффекты 3D. Они только отвлекают внимание читателя от сути вашего сообщения.

## ПРАВИЛО 4. ЛЕГКОЕ СРАВНЕНИЕ ДАННЫХ

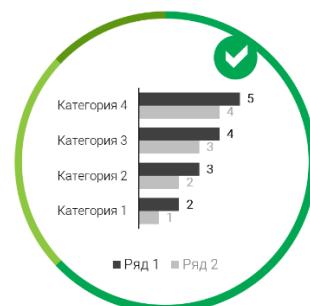
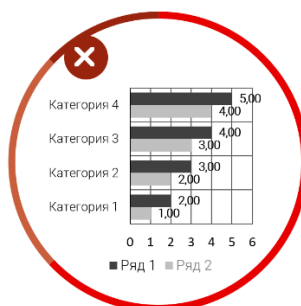
Визуализируйте данные так, чтобы их можно было легко сравнивать.



Одна из главных целей визуализации – удобное и наглядное сравнение двух и более показателей.

## ПРАВИЛО 5. МИНИМУМ ЭЛЕМЕНТОВ

На диаграмме должны быть только необходимые элементы.



Убирайте с ваших графиков и диаграмм все неинформативные элементы, оставляйте только необходимые.

## ПРАВИЛО 6. НЕ ПЕРЕГРУЖАЙТЕ ИНФОРМАЦИЕЙ

Следите, чтобы не было визуальной загроможденности.



Не пытайтесь уместить на одну диаграмму всю имеющуюся у вас информацию ради того, чтобы ваш график казался умным и значительным. Визуальный ряд не должен быть перегружен сложными и многоярусными диаграммами.

## ПРАВИЛО 7. ПОНЯТНЫЙ ФОРМАТ ЧИСЕЛ

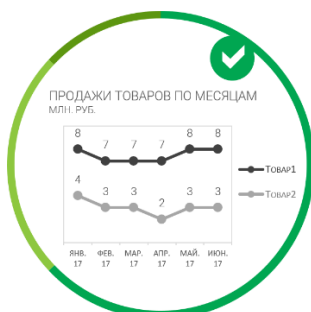
Числа должны быть с разделителями разрядов и без лишних знаков после запятой.



В больших числах всегда разделяйте разряды: 10 000 000, а не 10000000, иначе цифры становятся нечитабельными.

## ПРАВИЛО 8. НАЗВАНИЕ И ПОДПИСИ

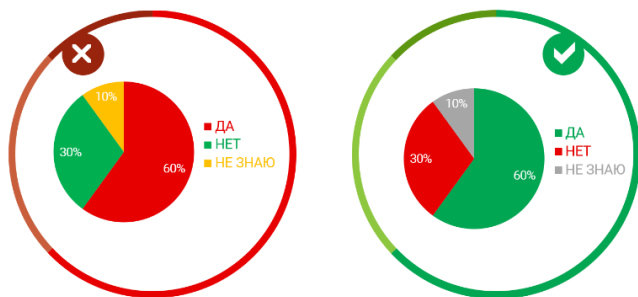
У диаграммы должно быть название и полная легенда.



Следите, чтобы всегда у вашей диаграммы было полное понятное название и все необходимые подписи, иначе появляется риск неверного истолкования.

## ПРАВИЛО 9. ОБЩЕПРИНЯТЫЕ ЦВЕТОВЫЕ РЕШЕНИЯ

Не нарушайте общепринятое использование того или иного цвета.

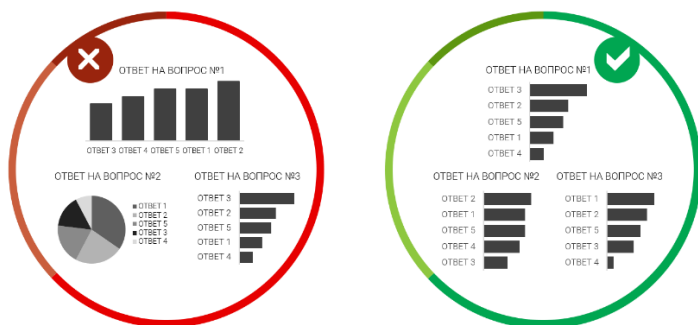


Есть несколько основных категорий, которые у нас всегда ассоциируются с определенным цветом:

1. положительные и отрицательные значения: зелёный и красный;
2. да/нет, согласен/не согласен: зелёный и красный;
3. мужчины и женщины: голубой и розовый;
4. прочее/другое/остальное/нет ответа/затрудняюсь ответить — серый цвет.

## ПРАВИЛО 10. МИНИМУМ ТИПОВ ДИАГРАММ

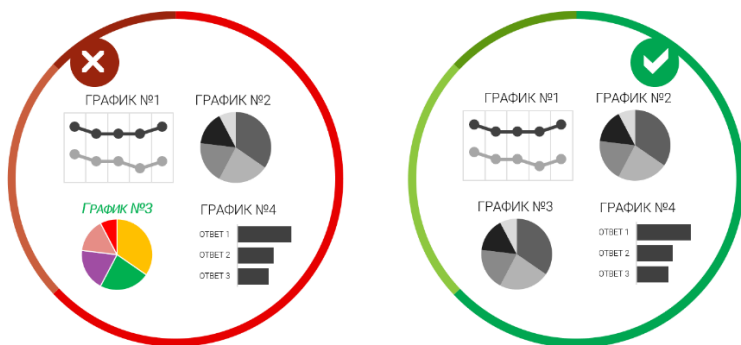
Используйте один вид диаграммы для однотипных данных.



Избегайте разнообразия ради разнообразия.

## ПРАВИЛО 11. ЕДИНАЯ ЦВЕТОВАЯ ПАЛИТРА

Придерживайтесь одной цветовой гаммы.



Визуальные элементы (графики, диаграммы, схемы) на протяжении всего исследования или отчета должны быть выполнены в одной цветовой гамме.

**Задание** (*Уровни 1, 2, 3*): воспользоваться указанными программами для создания собственных облаков слов. Найти как минимум 3 аналогичных ресурса.

## 2) Word Clouds

Free Word Cloud Generator

какого воплощается в музыкально-хореографических образах. В основе классического балетного спектакля лежит определённый сюжет, драматургический замысел, либретто, в XX веке появился «бессюжетный» балет, драматургия которого основана на развитии, заложенном в музыке. Основными видами танца в балете являются классический танец и характерный танец, к которому, начиная с XIX века, относятся народные и национальные танцы, переработанные для исполнения в балетном спектакле. Немаловажную роль играет пантомима, с помощью которой актёры передают чувства героев, их «разговор» между собой, суть происходящего, а во многих спектаклях — также и гротеск.

783 / 20,000

Тип источникаТекст

Создание облака слов

Вставить демонстрационный текстОчистить

ШрифтЛобстер

Цветовая палитра

Отображаемые слова60 / 60

Угол наклона текстаСлучайный

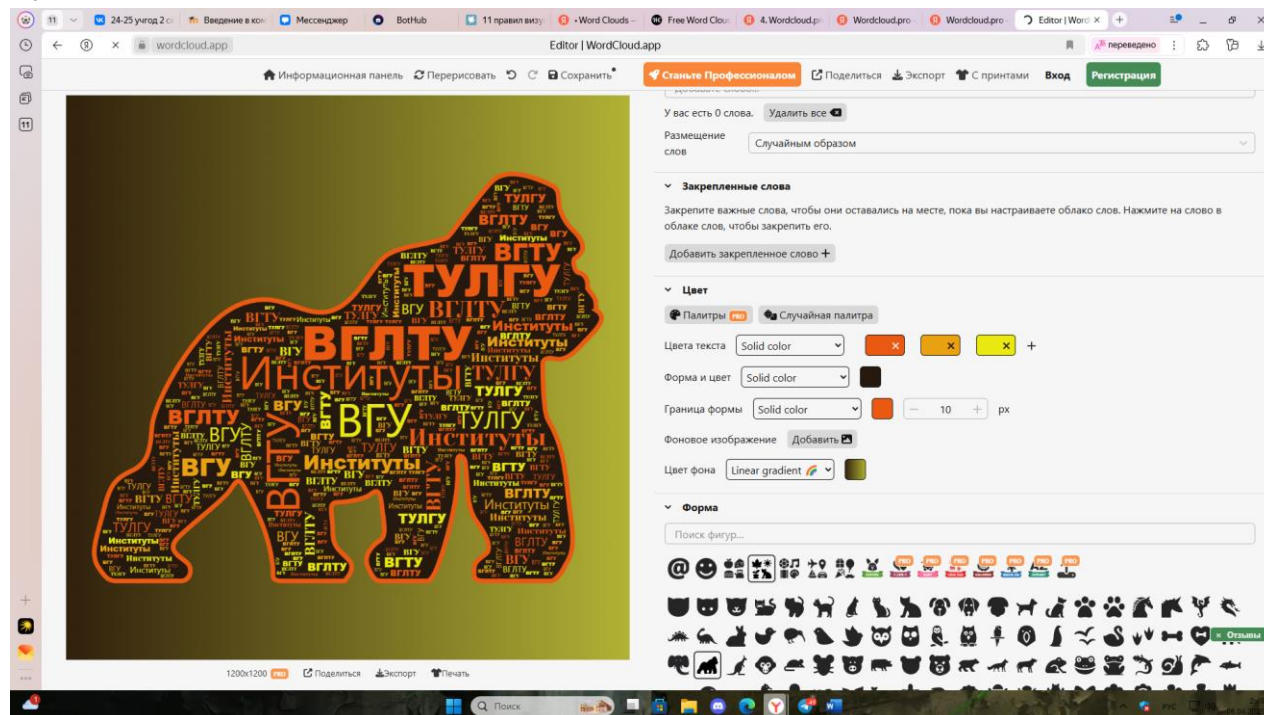
## 2) [WordItOut](#)



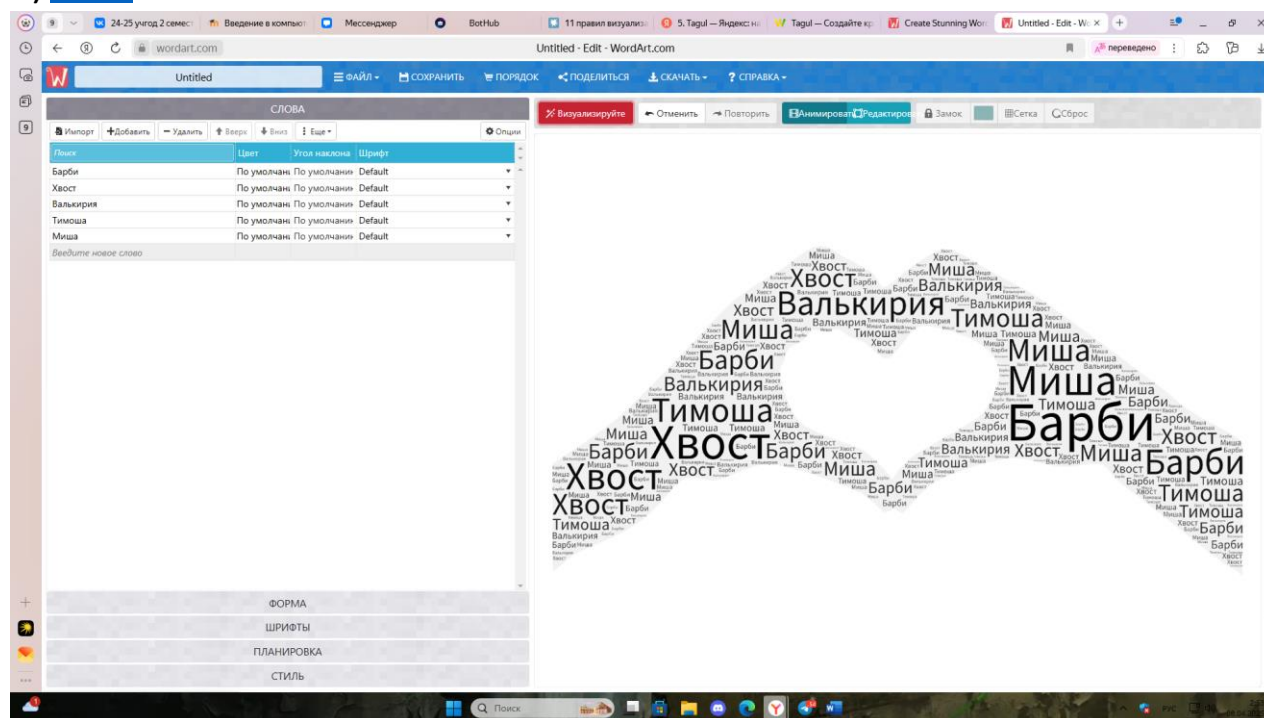
## 3) [Word Cloud by Jason Davies](#)



#### 4) [Wordcloud.pro](https://wordcloud.pro)



#### 5) [Tagul](https://tagul.com)



Похожие онлайн-сервисы для создания облаков тегов:

[WordArt](https://wordart.com)

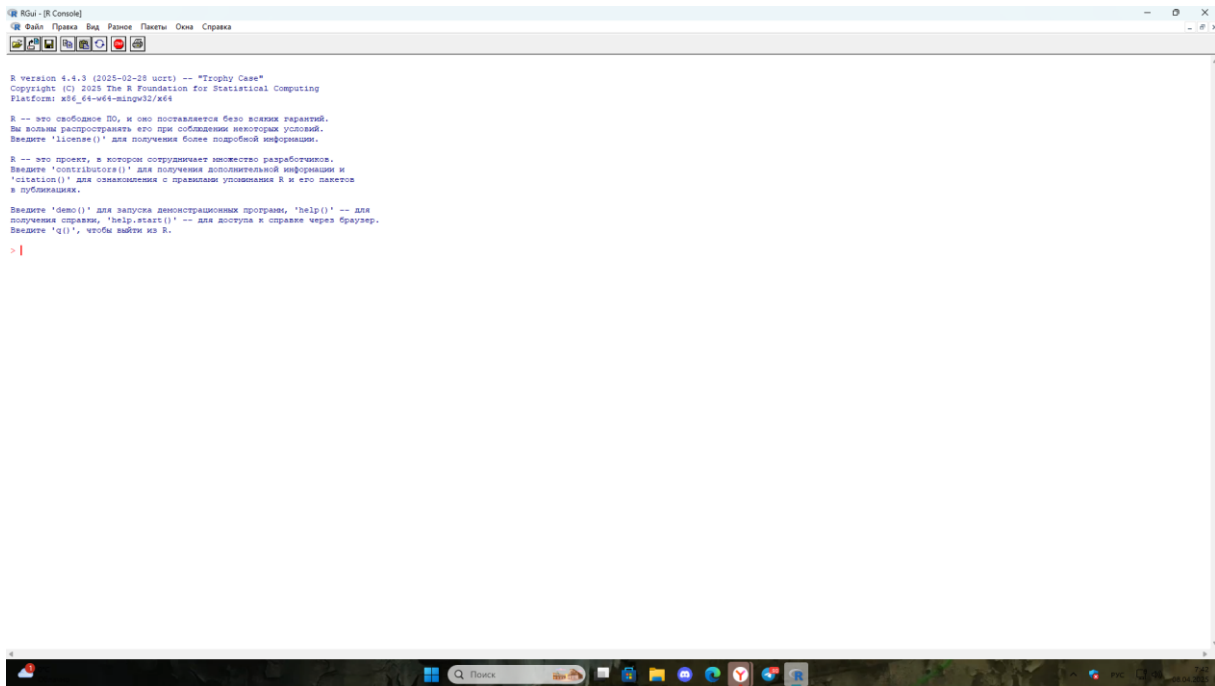
[Wordscloud.pythonanywhere.com](https://wordcloud.pythonanywhere.com)

[ImageChef](https://imagechef.com)

2)

## Стилеметрическое исследование:

### Установка пакета stylo:



```
RGui - [R Console]
Файл Правка Вид Панель Паquetes Окна Справка

R version 4.4.3 (2025-02-28 ucrt) -- "Trophy Case"
Copyright (c) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

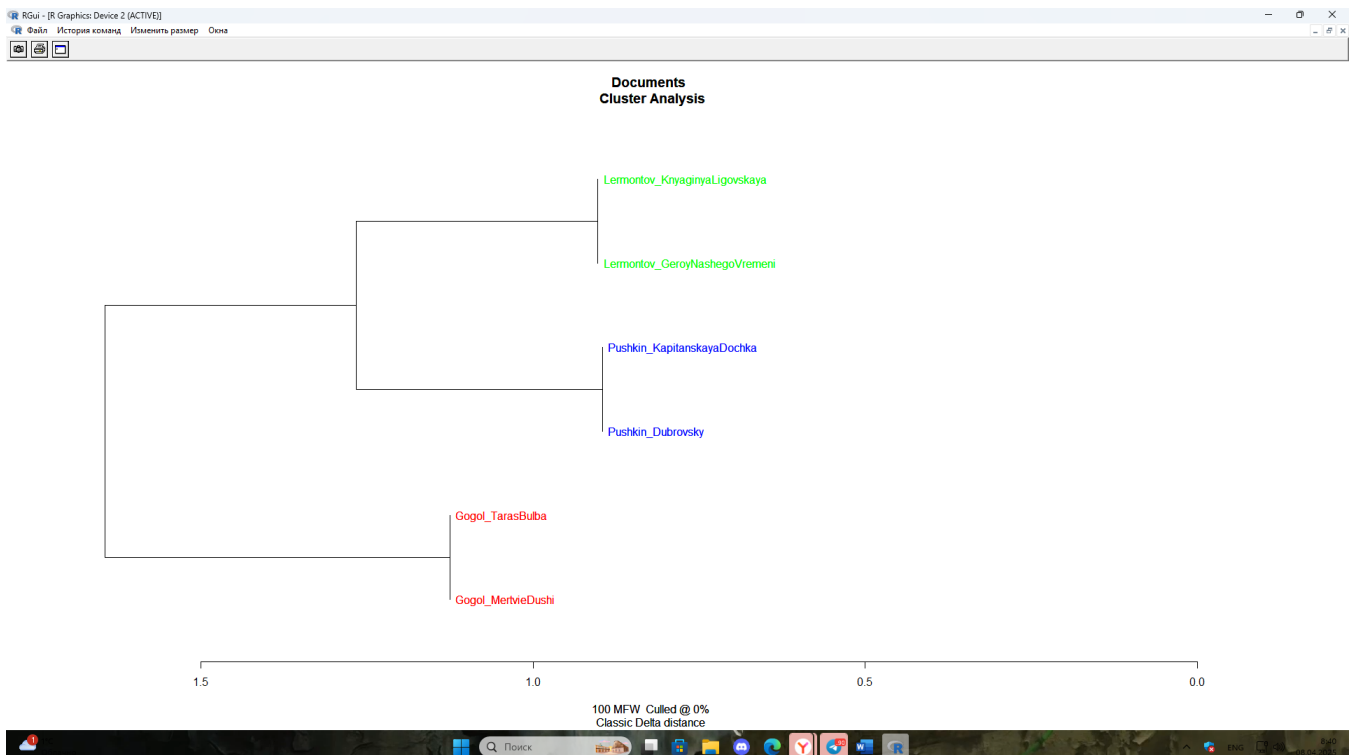
R -- это свободное ПО, и оно предоставляется без каких-либо гарантий.
Мы не можем распространять его при соблюдении некоторых условий.
Введите 'license()' для получения более подробной информации.

R -- это проект, в котором сотрудничает множество разработчиков.
Введите 'contributors()' для получения дополнительной информации и
'citation()' для ознакомления с правилами цитирования R и его пакетов
в публикациях.

Введите 'demo()' для запуска демонстрационных программ, 'help()' -- для
получения справки, 'help.start()' -- для доступа к справке через браузер.
Введите 'q()', чтобы выйти из R.

> |
```

### Результаты кластеризации:



Вывод: кластеризация прошла успешно и тексты разделились по своим авторам



3)

Ресурсы:

<http://www.rusf.ru/books/analysis/> — Хмелев Д. В. Лингвоанализатор — авторский сайт, посвященный вопросу определения авторства

<http://attribution.corneille-moliere.com/?p=page19&m=main&l=rus> — информационный ресурс посвящен проблеме определения авторства анонимных и псевдонимных произведений

<http://postnauka.ru/faq/11053>

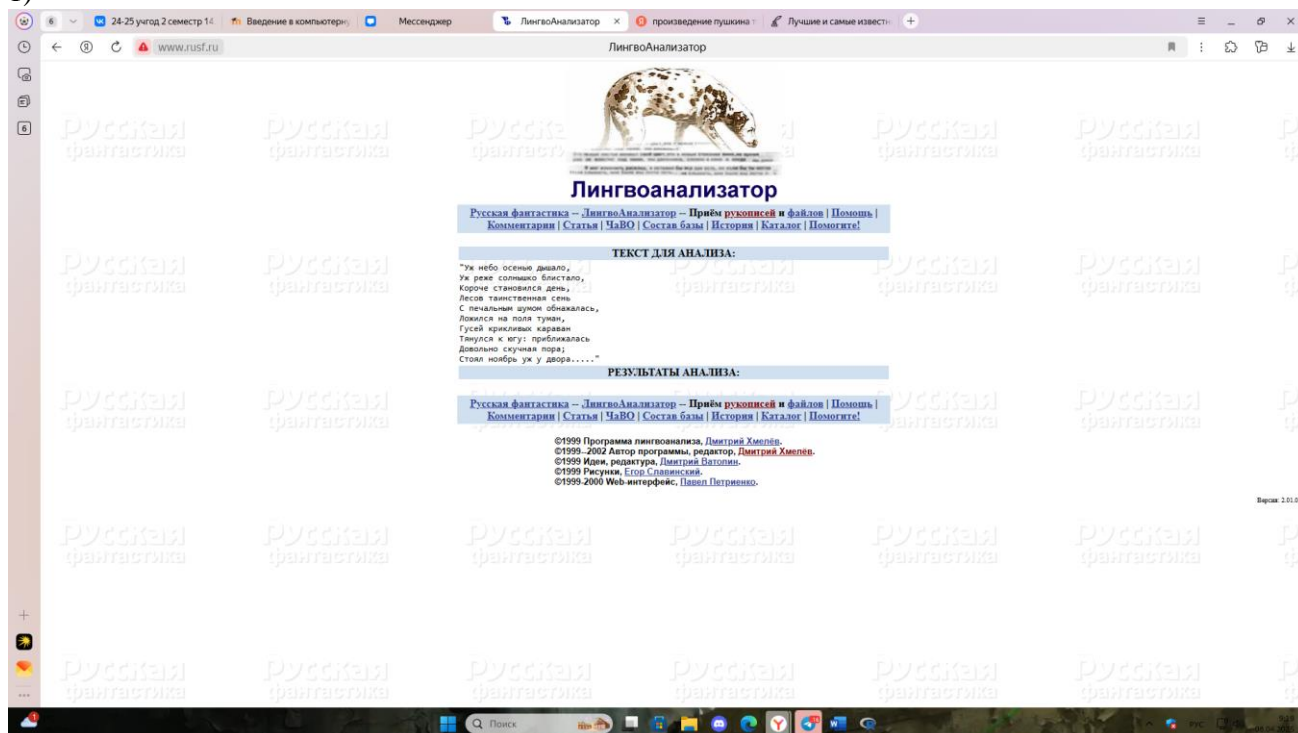
<http://starling.rinet.ru/Texts/textsr.htm>

<http://dict.ruslang.ru/freq.php> — новейший академический частотный словарь русского языка

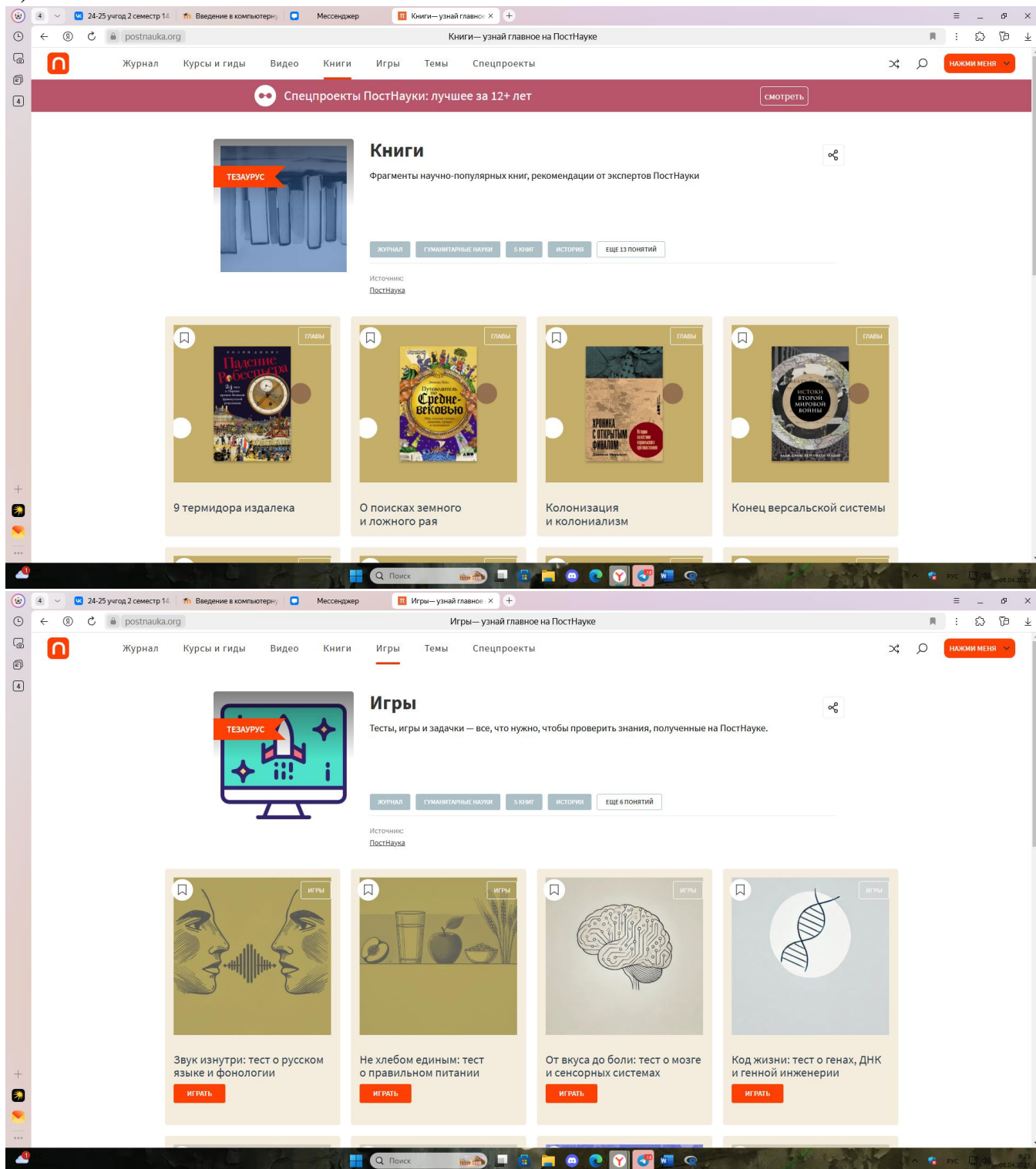
<http://www.artint.ru/projects/frqlist.php> — частотный словарь одного из авторов предыдущего словаря

<http://morpher.ru/Russian/Stats.aspx> — частотный словарь, созданный программистом для сравнения с предыдущими двумя

1)



2)



3)

24-25 учгод 2 семестр 14 Введение в компьютерн Мессенджер 32 Вавилонская башня 32 Вавилонская башня

Word starlingdb.org Вавилонская башня

Страница С. А. Старостина English

## Статьи и книги: электронная библиотека Вавилонской башни

[Все тексты выложены в формате PDF за исключением специально оговоренных случаев]

### Работы общего и методологического характера

С. А. Старостин. *Сравнительно-историческое языкознание и лексикостатистика*. // Впервые опубликовано в: *Лингвистическая реконструкция и древнейшая история Востока*. М., Наука, 1989, стр. 3-39. Перепечатано в: Старостин С. А. Труды по языкознанию. М., Языки славянских культур, 2007, стр. 407-447. [по-русски](#)

И. И. Пейрос. *Макросемы: можно ли обнаружить ошибку?* // Впервые опубликовано в: *Indo-European, Nostratic, and Beyond: Festschrift for Vitaly V. Shevoroshkin*. Ed. by Iren Hegedus, Peter A. Michalove and Alexis Manaster Ramer. JIES Monograph No. 22. Institute for the Study of Man, Washington DC, 1997, стр. 265-292. [по-английски](#)

И. И. Пейрос. *Эволюция языковых семей, языковая история и генетическая классификация*. // Впервые опубликовано в: *Historical Linguistics & Lexicostatistics*. Ed. by Vitaly Shevoroshkin & Paul J. Sidwell. Melbourne, 1999, стр. 257-305. [по-английски](#)

С. А. Старостин. *О доказательстве языкового родства*. // Впервые опубликовано в: Типология и теория языка (к 60-летию А. Е. Кибрика). М., Языки русской культуры, 1999, стр. 57-69. Перепечатано в: Старостин С. А. Труды по языкознанию. М., Языки славянских культур, 2007, стр. 779-793. [по-русски](#)

С. А. Старостин. *Методология дальнего сравнения*. // Впервые опубликовано в: *Historical Linguistics & Lexicostatistics*. Ed. by Vitaly Shevoroshkin & Paul J. Sidwell. Melbourne, 1999, pp. 61-66. [по-английски](#)

Мерритт Рулен. *Описание японской типологической базы данных*. // Рукопись, 2004. [по-английски](#)

А. В. Дыбо, Г. С. Старостин. *В защиту сравнительного метода*. // Впервые опубликовано в: *Аспекты компаративистики* 3 М., изд. РГГУ, 2008, стр. 109-258. [по-английски](#)

### Индоевропейская семья

В. А. Дыбо. *Балто-славянская акцентология и закон Винтера (часть)*. // Впервые опубликовано в: *Studia Linguarum*, в. 3 М., изд. РГГУ, 2002, стр. 295-515. [по-английски](#)

В. А. Дыбо. *Балтийская сравнительно-историческая и литовская историческая акцентология*. // Впервые опубликовано в: *Аспекты компаративистики* 1 М., изд. РГГУ, 2005, стр. 177-213. [по-русски](#)

В. А. Дыбо. *Рефлексы индоевропейских словесных долгов: племена и носители в кельто-италийских языках (в защиту концепции Ж. Пейрона)*. // Впервые опубликовано в: *Аспекты компаративистики* 2 М., изд. РГГУ, 2007, стр. 11-27. [по-русски](#)

В. А. Дыбо. *Германское сокращение индоевропейских долгов, германский "Vergleichung" (закон Хольцмана) и балто-славянская акцентология*. // Впервые опубликовано в: *Аспекты компаративистики* 3 М., изд. РГГУ, 2008, стр. 537-608. [по-русски](#)

### Алтайская семья

С. А. Старостин. *Алтайская проблема и происхождение японского языка [МОНОГРАФИЯ]*. // М., Наука, 1991. [по-русски](#)

А. В. Дыбо. *Древнейшие язычные в транскориксе*. // Впервые опубликовано в: *Аспекты компаративистики* 1 М., изд. РГГУ, 2005, стр. 49-82. [по-русски](#)

О. А. Мудрак. *Заметки о языке и культуре бурайских булар*. // Впервые опубликовано в: *Аспекты компаративистики* 1 М., изд. РГГУ, 2005, стр. 83-106. [по-русски](#)

А. В. Дыбо. *Реконструкция правдоуказного строения*. // Впервые опубликовано в: *Аспекты компаративистики* 2 М., изд. РГГУ, 2007, стр. 259-280. [по-русски](#)

О. А. Мудрак. *Развитие сочетаний c1 в праалтайском языке*. // Впервые опубликовано в: *Аспекты компаративистики* 2 М., изд. РГГУ, 2007, стр. 337-352. [по-русски](#)

### Эскимосская семья

О. А. Мудрак. *Чукотско-кавказская и эскимосская аллотропология и некоторые алтайско-эскимосские этимологии, обнаруживаемые в списке Суодея*. // Впервые опубликовано в: *Аспекты компаративистики* 3 М., изд. РГГУ, 2008, стр. 297-336. [по-английски](#)

### Дравидийская семья

Г. С. Старостин. *Аллотропные согласные в дравидийском: один или больше?* // Впервые опубликовано в: *Proceedings of the International Conference on South Asian Languages (July 1-4, 1997)*. Moscow: Institute of Asian and African Studies, 1998, pp. 183-194. [по-английски](#)

Г. С. Старостин. *Еще раз к вопросу о личных местоимениях в дравидийском языке*. // Впервые опубликовано в: *Аспекты компаративистики* 2 М., изд. РГГУ, 2007, стр. 101-150. [по-русски](#)

### Афразийская макросемья

О. А. Мудрак. *Еще раз о аллотропологии и сравнительном методе: омонимии и афразийский*. // Впервые опубликовано в: *Аспекты компаративистики* 1 М., изд. РГГУ, 2005, стр. 339-408. [по-английски](#)

О. А. Мудрак. *Расширители корня и образование новых корней в семитском и афразийском*. // Впервые опубликовано в: *Proceedings of the Barcelona Symposium on comparative Semitic, 19-20/11/2004*. Aula Orientalis 23/1-2, 2005, pp. 83-129. [по-английски](#)

### Сино-кавказская макросемья

С. А. Старостин. *Конечные согласные в сино-кавказских языках*. // Впервые опубликовано в: *Journal of Chinese Linguistics*, vol. 24, 1996, стр. 281-311. Перепечатано в: С. А. Старостин. Труды по языкознанию. М., Языки славянских культур, 2007, стр. 633-659. [по-английски](#)

С. А. Старостин. *Сино-кавказский (сравнительная фонология)*. // Рукопись, 2004-05. [по-английски](#)

С. А. Старостин. *Сино-кавказский (глоссарий)*. // Рукопись, 2004-05. [по-английски](#)

Джон Бенгтсон. *Материалы к сравнительной грамматике дине-кавказских (сино-кавказских) языков*. // Впервые опубликовано в: *Аспекты компаративистики* 3 М., изд. РГГУ, 2008, стр. 45-118. [по-английски](#)

4)

24-25 учгод 2 семестр 14 Введение в компьютерн Мессенджер 32 Вавилонская башня 32 Вавилонская башня

Word dict.ruslang.ru Новый частотный словарь русской лексики

## СЛОВАРИ НА ОСНОВЕ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА

О. Н. Ляшевская, С. А. Шаров  
НОВЫЙ ЧАСТОТНЫЙ СЛОВАРЬ РУССКОЙ ЛЕКСИКИ

Электронная версия издания:  
О. Н. Ляшевская, С. А. Шаров, Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.

Как пользоваться словарем

Введение к Новому частотному словарю  
CSV-версия словаря

### I. Общая лексика

- Алфавитный список лемм
- Частотный список лемм
- Распределение лемм по функциональным стилям:
  - Частотный словарь художественной литературы
  - Словарь значимой лексики художественной литературы
  - Частотный словарь публицистики
  - Словарь значимой фактивно-новостной лексики
  - Частотный словарь другой нехудожественной литературы
  - Частотный словарь значимой другой нехудожественной литературы
  - Частотный словарь живой устной речи
  - Словарь значимой лексики живой устной речи
  - Алфавитный список словоформ

### II. Общая лексика: части речи

- Частотный список имен существительных
- Частотный список глаголов
- Частотный список имен прилагательных
- Частотный список наречий и предикативов
- Частотный список местоимений (местоимения-существительные, прилагательные, наречия, предикативы)
- Частотный список числительных
- Частотный список лемм служебных частей речи

### III. Вспомогательные таблицы

- Данные о частотности частеречных классов
- Частотность букв русского алфавита
- Частотность двубуквенных сочетаний

### IV. Имена собственные и аббревиатуры

- Алфавитный список собственных имен и аббревиатур

5)

**Частотный словарь**

C.A. Шаров

**Вторая версия частотного списка**

На этой странице Вы можете получить списки наиболее частотных слов русского языка. До настоящего времени Частотный словарь русского языка под ред. Л.Н.Засориной (1977) чаще всего использовался в качестве источника информации о частоте русских слов. Однако корпус, на основе которого была подсчитана частота слов в этом словаре, по современным стандартам очень мал (около миллиона слов). Кроме того, список существенно устарел: он соответствует частоте использования слов в период с 20-х до 60-х годов. В результате корпус включает большое число идеологических источников, например, произведения Ленина и Калинина. Материалы 22 и 23 съездов КПСС и советские газеты. Слова советской и ленинской эпохи входят в левую часть русских слов, наряду со служебными словами (они встречаются чаще слов *аде*, *здесь*, *яши*), слова *партия*, *революция*, *коммунистический* встречается чаще чем *назад*, *около*, *лучше* и т.д. Наконец, список слов из словаря Засориной не существует в электронном виде.

Список слов, доступный с этой страницы, содержит примерно 35000 слов с частотой большей 1 ipm (вхождений на миллион слов, instances per million words). Имеется также более короткий список из 5000 наиболее частотных русских слов. Списки используют кодировку кириллицы utf8 и упакованы утилитой WinZip (пользователи Linux или Mac могут использовать Stuffit для распаковки).

Структура списков соответствует формату лемматизированных списков из British National Corpus (BNC), созданных Адамом Килгарифом, а именно: порядковый номер, частота (ipm), лемма, часть речи (классификация BNC).

**Слова с частотой больше 1 ipm**

- [lemma.al.zip](#) - леммы, отсортированные в алфавитном порядке
- [lemma.num.zip](#) - леммы, отсортированные по частоте
- [words.num.zip](#) - словоформы, отсортированные по частоте

**Список 5000 наиболее частых слов**

- [5000lemma.al.zip](#) - леммы, отсортированные в алфавитном порядке
- [5000lemma.num.zip](#) - леммы, отсортированные по частоте

**Некоторые статистические данные об использовании русских слов**

- Средняя длина слова 5.28 символа.
- Средняя длина предложения 10.38 слов.
- 1000 наиболее частотных лемм покрывают 64.0708% текста.
- 2000 наиболее частотных лемм покрывают 71.9521% текста.
- 3000 наиболее частотных лемм покрывают 76.5104% текста.
- 5000 наиболее частотных лемм покрывают 82.0604% текста.

Более полная информация о соответствии частотой слова и покрытием корпуса находится [здесь](#).

Список построен на основе представительного корпуса современного русского языка. Он включает в себя подборку современной прозы, политических мемуаров, современных газет и научно-популярной литературы (около 40 миллионов слов, проза составляет примерно чуть больше половины объема). Все тексты корпуса были написаны на русском в промежутке между 1970 и 2002, большинство между 1980 и 1995, газетный корпус 1997-1999 (корпус основан на текстах из Библиотеки Мошкова и корпуса современной публицистики А.В.Баранова).

Хорошо известно, что большие тексты представляют проблему для составления частотных списков, поскольку относительно длинный текст может содержать большое количество вхождений некоторого редкого слова, что существенно увеличит его частоту в итоговом списке. Например, корпус, использованный для составления данного списка, содержит вариацию на тему Толкиенивского "Повелителя Колец" (автор Ник Терунов). Несмотря на то, что длина этого романа составляет 250 тыс. слов, менее одного процента всего корпуса, частота использования слова *хоббит* в этом романе ставит его в первую тысячу русских слов, если частоту считать по всем текстам без ограничений на их длину. По этой причине частотные списки были составлены при условии, что выборка из больших текстов ограничена 10 тыс. слов, и выборка из текстов одного автора составляет менее 100 тыс. слов. В результате подмножество полного корпуса, использованное при подсчете частоты, составляет около 16 миллионов слов.

Распределение слов в текстах далеко от равномерного. Некоторые слова (например, *предлог*), встречаются во многих текстах с вполне предсказуемой частотой. Частота других (например, *мостомойки* или *ментальных глаголов*) существенно зависит от автора или жанра текста, в то время как многие слова относятся к "зарезаным", если это слово (например, или собственное, обозначение человека по званию или должности или технический термин) встретилось в тексте один раз, весьма вероятно, что оно повторится там еще много раз, таким образом, существенно повышая его частоту в документе. Существуют разные способы измерения такой вариации (Church, K. and Gale, W. (1995) *Poisson Mixtures, Journal of Natural Language Engineering*, 1.2). Простейший способ для оценки поведения слова, посчитать коэффициент вариации, который вычисляется как среднеквадратичное отклонение, поделенное на среднее значение. Среднеквадратичное отклонение дает абсолютное значение вариации набора данных (оно увеличивается для слов с большей средней частотой), в то время как коэффициент вариации позволяет сравнить распределение слов с неравной средней частотой. Значения отклонений для 5000 наиболее частотных слов можно посмотреть [здесь](#). Структура файла:

лемма, средняя частота (ipm), число текстов, в которых это слово встречается, среднеквадратичное отклонение частоты по всем текстам, коэффициент вариации, дисперсия.

Корпус, средства для работы с ним, а также параллельный англо-русский корпус (выравнивание на основе предложений) описаны, в частности, в следующей публикации автора:

6)

**Морфер**

+7 (925) 336 9960  
[sales@morpher.ru](mailto:sales@morpher.ru)

О программе | Продукты | Демо | Русский язык | Связаться с нами | [Мой Морфер](#)

**Частотный словарь полутора гигабайт русских текстов**

- Основные результаты
  - 300 наиболее часто встречающихся словоформ
  - Статистика корпуса текстов
  - Динамика роста размера словаря
- Детали реализации на C++
  - Требования к программному и аппаратному обеспечению
  - Производительность
  - Переносимость
  - Скачать
- Ссылки

Сентябрь 2002

У меня давно было желание провести статистическое исследование достаточно представительной выборки русских слов, чтобы найти среди них самые часто употребляемые, самые длинные, самые средние... в общем, все самые-самые.

И вот, когда у меня в руках оказался диск «Библиотека в кармане» (выпуск 4), я понял, что моя мечта близка к осуществлению. Написание программы заняло совсем немного времени, и теперь я с удовольствием представляю вам результаты этого эксперимента.

**Основные результаты**

Было просканировано около 1,5Гб (более полумиллиона печатных страниц) русских текстов различной тематики: классика, детективы, фантастика, специальная литература, словари и энциклопедии. Общая длина всех текстов составила более двухсот миллионов слов. (Точнее, словоформ. Здесь и далее речь будет идти только о словоформах.) Среди этих двухсот миллионов уникальных оказались всего лишь чуть более двух миллионов словоформ, причем 47% из них встретились только один раз. (Любители точных цифр найдут их в результатах одного из прогонов [здесь](#).)

Наиболее часто употребляемыми оказались, как и следовало ожидать, служебные слова –

## Похожие ресурсы

### 1. Научные публикации и статьи по анализу текстов:

- Google Scholar (<https://scholar.google.com/>) - ресурс для поиска научных статей и публикаций по различным темам, включая лингвистику и анализ авторства.

### 2. Лингвистические и литературоведческие ресурсы:

- Национальный корпус русского языка (<http://www.ruscorpora.ru/>) - предоставляет доступ к большому количеству текстов на русском языке и инструментам для их анализа.

- Всероссийская научная библиотека (<http://www.vdam.ru/>) - содержит различные цифровые коллекции работ и исследований в области лингвистики.

### 3. Инструменты для стилистического и лексического анализа:

- Text Analysis Online (<https://text-analysis-online.com/>) - онлайн-инструмент для анализа текстов, включая частотный анализ слов.

- RusLang (<http://ruslang.ru/>) - сайт с различными инструментами для лексического и морфологического анализа текста на русском языке.

### 4. Сайты с частотными словарями и лексикографическими данными:

- Частотный словарь современного русского языка ([http://www.ruscorpora.ru/from\\_corpora](http://www.ruscorpora.ru/from_corpora)) - архив частотных словарей и инструментов для их анализа.

- Словари и грамматики для русского языка (<https://www.russiangrammar.com/>) - содержит частотные словари и инструменты по морфологии..

JGAAP    Java Graphical Authorship Attribution Program    Алгоритмы AT, визуализация графов, анализ n-grams <http://www.jgaap.sourceforge.net/>

Stylometry with R    Пакет для анализа текста в R, позволяющий использовать разные методы стилистического анализа. Многократные методы анализа, включая LDA и n-grams <https://cran.r-project.org/web/packages/stylometry/>

Authorship Attribution Tool    Онлайн-сервис для анализа текстов и оценки вероятного автора  
Автоматическое определение авторства, графический интерфейс <http://www.authorship.org/>

Text Mining Toolkit    Набор Python-библиотек для анализа текстов с функциями атрибуции NLP, частотный анализ, сравнительный анализ    <https://www.nltk.org/>

TidyText    Расширение для анализа текстов в R через tidyverse    Экосистема анализов, частотный анализ, визуализация <https://www.tidytextmining.com/>

KSPM    Методы машинного обучения для определения авторства текстов    Использует алгоритмы машинного обучения <https://github.com/swett/kspm>

Lingua    Библиотека для работы с языковыми данными в Python    Обработка текста, анализ частот и структуры    <https://github.com/plutonium/lingua>

Syllable Count Analyzer    Инструмент для анализа текста с учетом структуры слова    Анализ слогов для стили написания    <https://www.syllablecount.com/>

1. JGAAP - это мощная программная платформа, предоставляющая широкий набор инструментов для анализа авторства текстов и визуализации результатов.

2. Stylometry with R - специфический пакет для языка R, который включает разнообразные алгоритмы стилистического анализа.

3. Authorship Attribution Tool - онлайн-сервис, который позволяет загружать тексты и получать результаты анализа на основе различных алгоритмов.

4. Text Mining Toolkit - предоставляет доступ к инструментам обработки текстов и анализа авторства в среде Python.

5. TidyText - библиотека для анализа текстов в R, которая интегрируется с другими пакетами tidyverse для эффективной обработки данных.

6. KSPM - фреймворк для выполнения анализа авторства с помощью машинного обучения, предоставляющий множество алгоритмов.

7. Lingua - библиотека для работы с языковыми данными и выполнения различных видов анализа текстов.

8. Syllable Count Analyzer - инструмент для анализа текстов с точки зрения стилистики, focusing on syllable counts and readability.