**Уровень 1.**

**Задание 1.**

**1. (Уровень 1, 2, 3)**

Диаграмма – графическое представление данных, позволяющее быстро оценить соотношение нескольких величин. Представляет собой геометрическое символьное изображение информации с применением различных приёмов техники визуализации.

Диаграммы-линии или графики – это тип диаграмм, на которых полученные данные изображаются в виде точек, соединённых прямыми линиями.

Классическими диаграммами являются столбчатые и линейные (полосовые) диаграммы. Также они называются гистограммами. Столбчатые диаграммы в основном используются для наглядного сравнения полученных статистических данных или для анализа их изменения за определённый промежуток времени. Разновидностями столбчатых диаграмм являются линейные (полосовые) диаграммы. Они отличаются горизонтальным расположением столбиков. Столбчатые и линейные диаграммы взаимозаменяемы, рассматриваемые в них статистические показатели могут быть представлены как вертикальными, так и горизонтальными столбиками.

Круговые (секторные) диаграммы

Достаточно распространённым способом графического изображения структуры статистических совокупностей является секторная диаграмма, так как идея целого очень наглядно выражается кругом, который представляет всю совокупность.

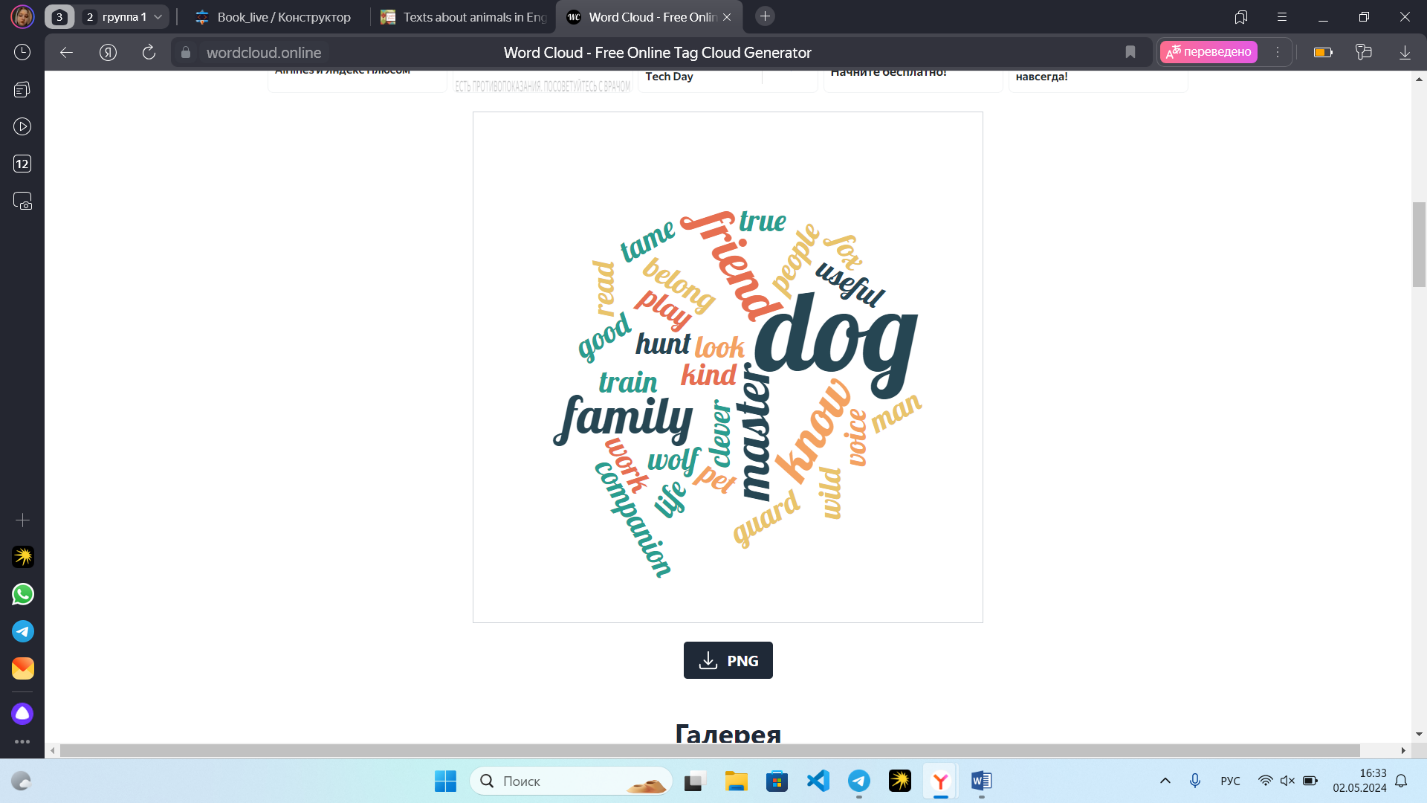
Радиальные (сетчатые) диаграммы

В отличие от линейных диаграмм, в радиальных или сетчатых диаграммах более двух осей. По каждой из них производится отсчёт от начала координат, находящегося в центре.

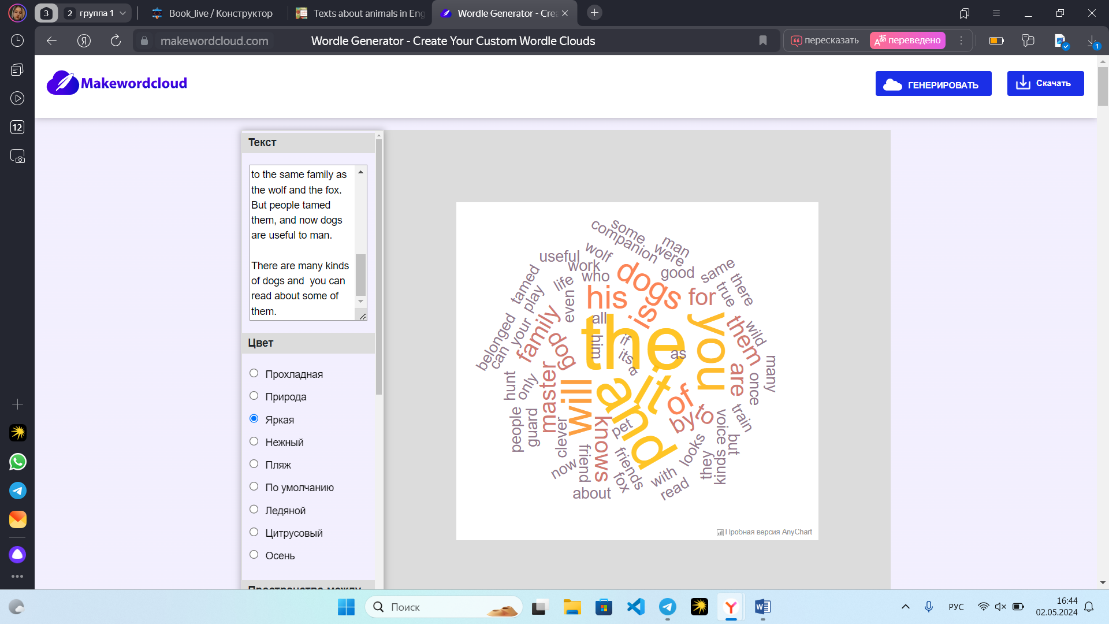
Таблица (из лат. tabula «доска») – способ передачи содержания, заключающийся в организации структуры данных, в которой отдельные элементы помещены в ячейки, каждой из которых сопоставлена пара значений – номер строки и номер колонки.

**2. (Уровень 1, 2, 3)**

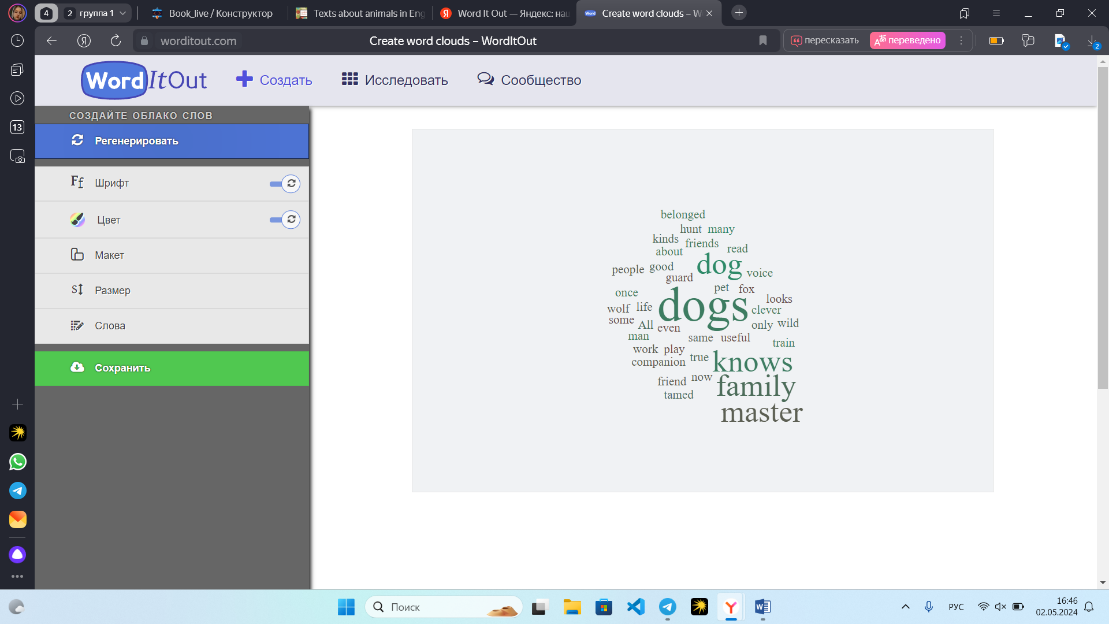
- [Word Clouds](https://www.wordclouds.com/)



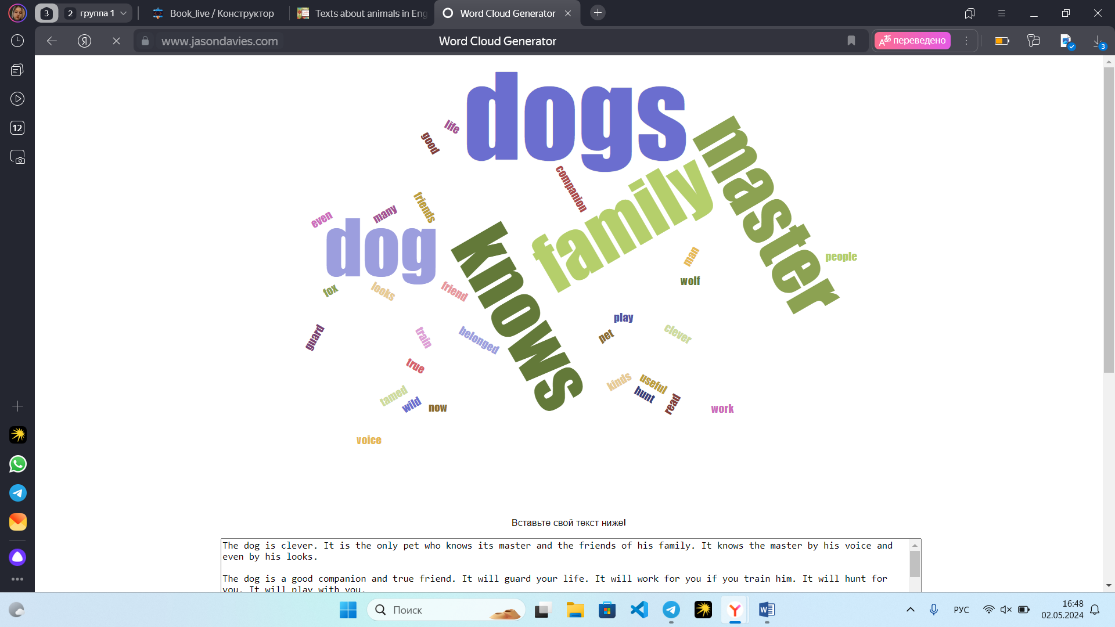
**-** [Wordle](http://www.wordle.net/)



- [Word It Out](https://worditout.com/)



- [Word Cloud by Jason Davies](https://www.jasondavies.com/wordcloud/)

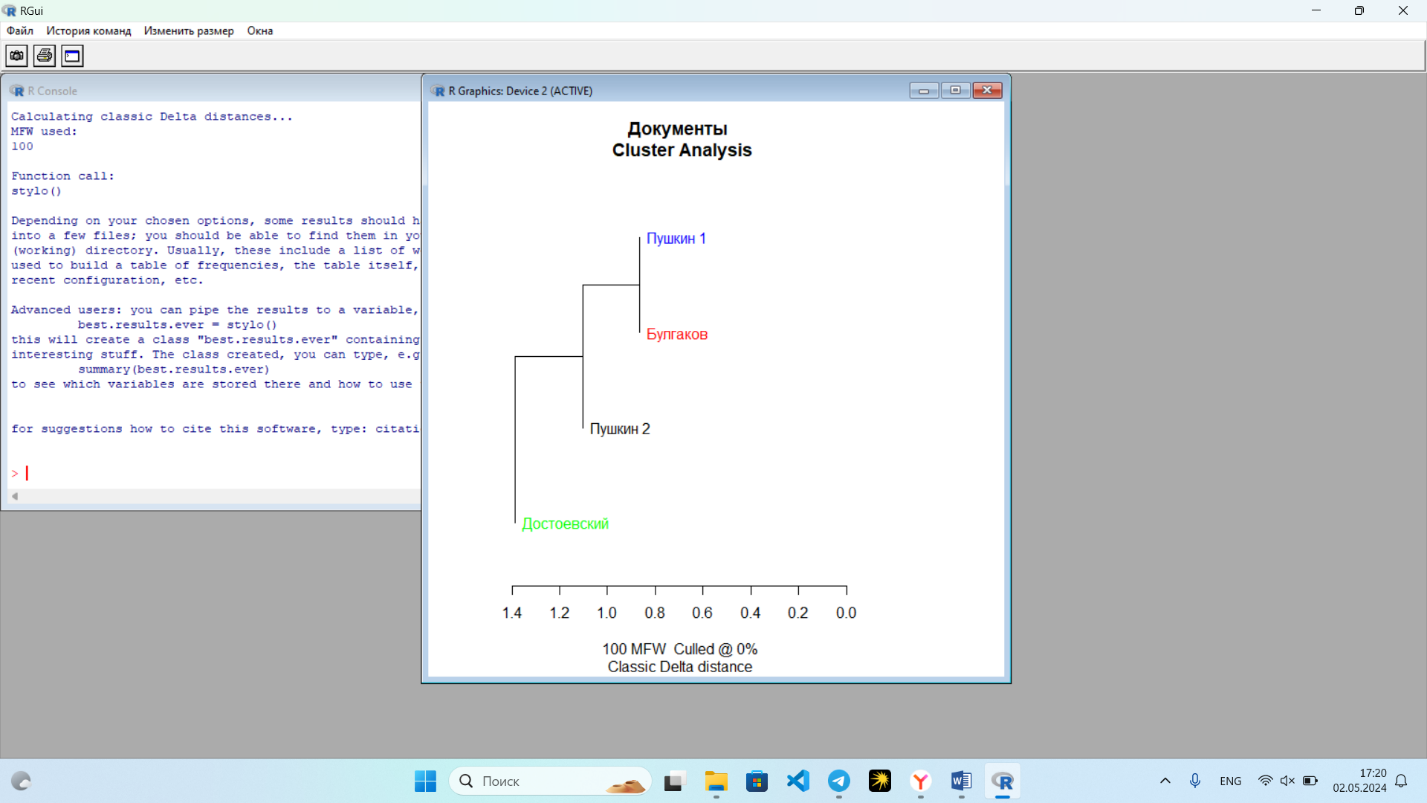


- Wordcloud.pro https:// wordcloud.pro/ru



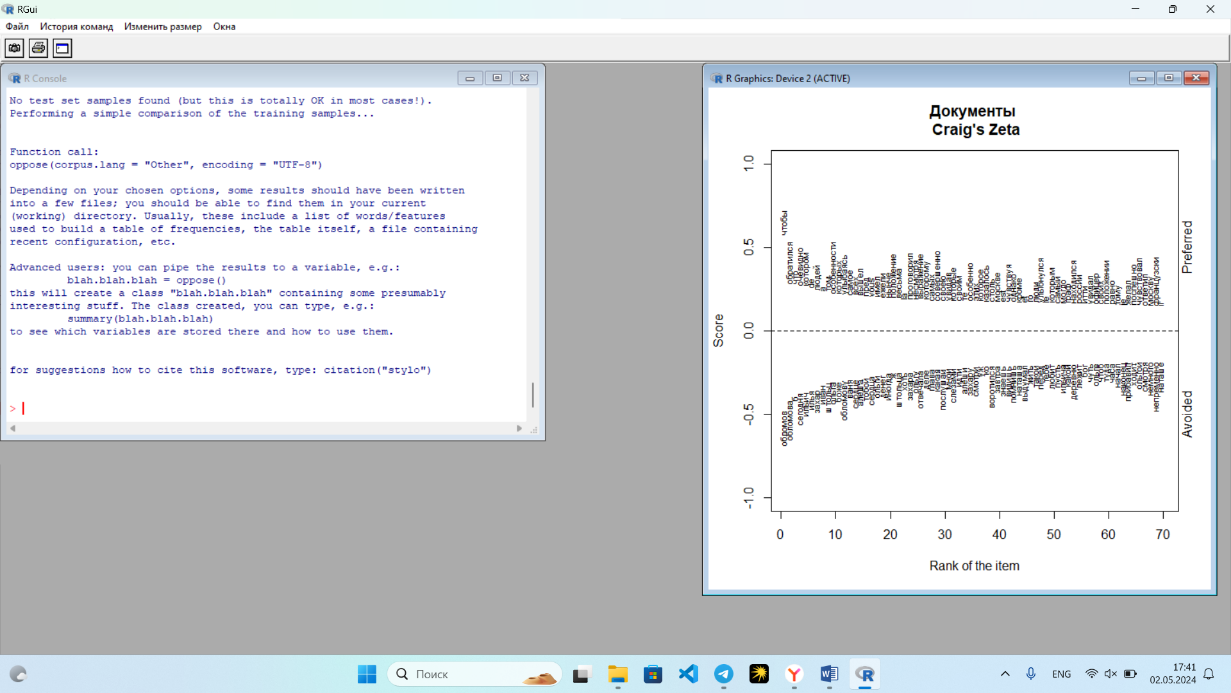
**Задание 2.**

**1. (Уровень 1, 2, 3)**

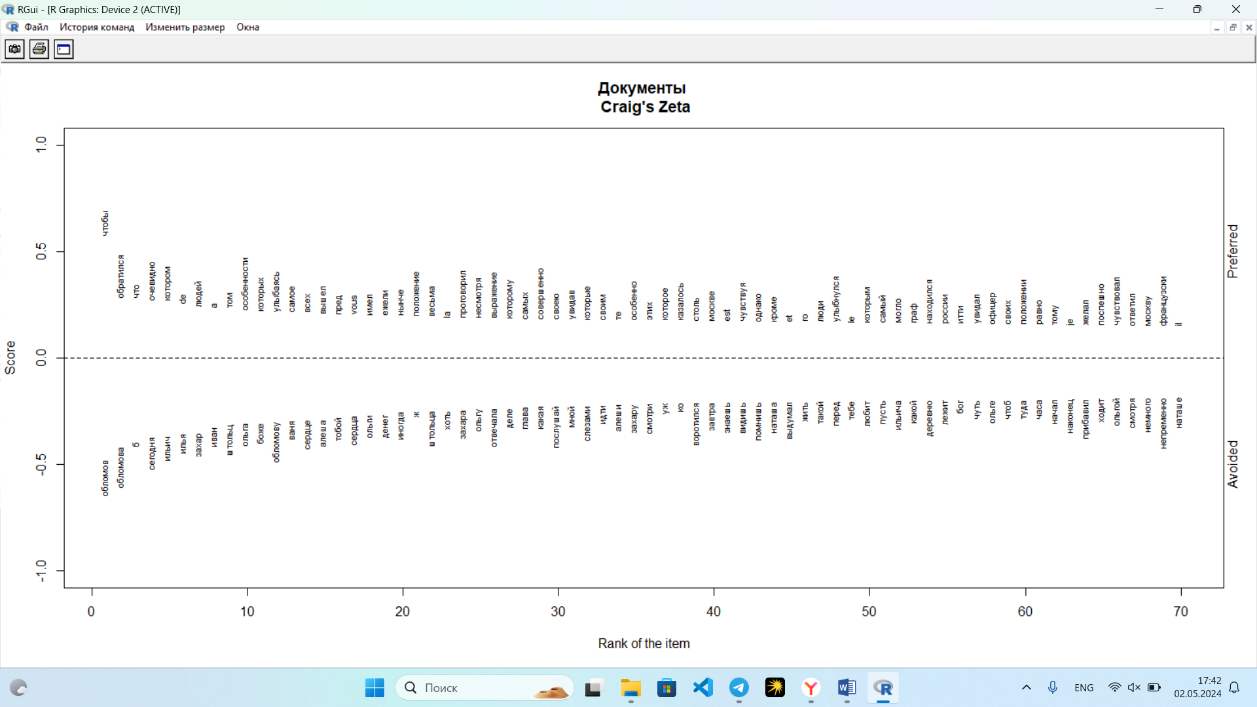
По графику видно, что стили текстов похожи, т. к. авторы стоят рядом друг с другом.

**2. (Уровень 2, 3)**



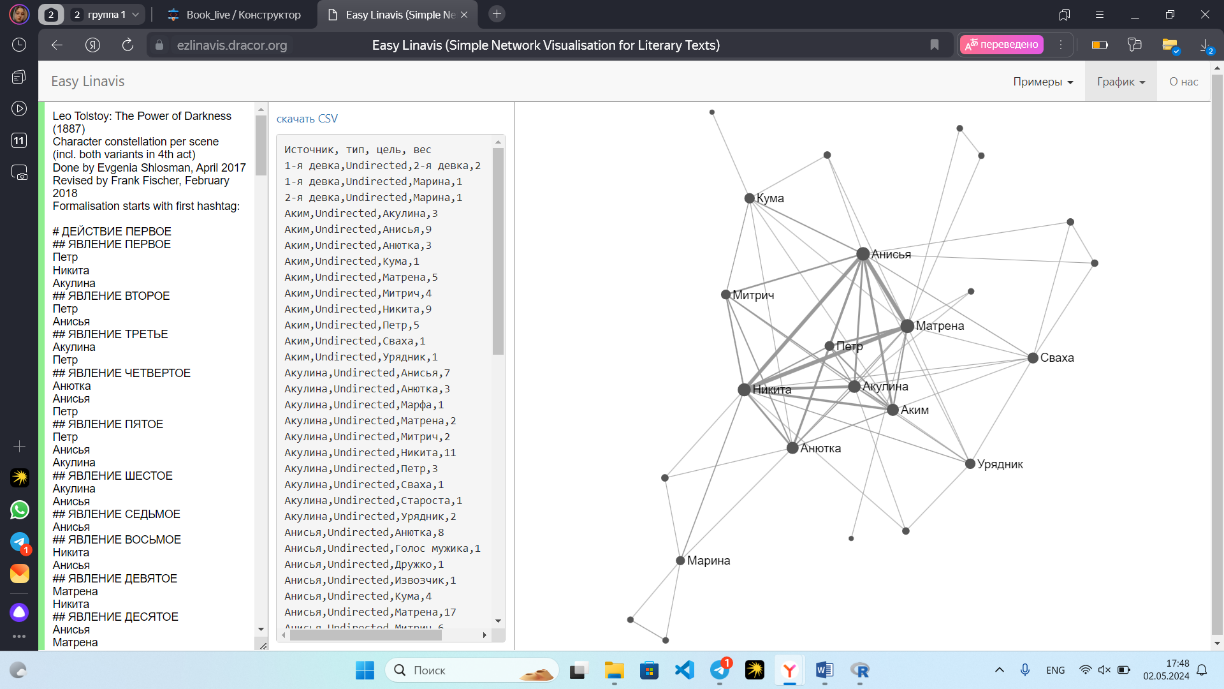


Тут видно какие слова предпочитают авторы, а каких слова избегают.



**Задание 3. (Уровень 1,2,3)**

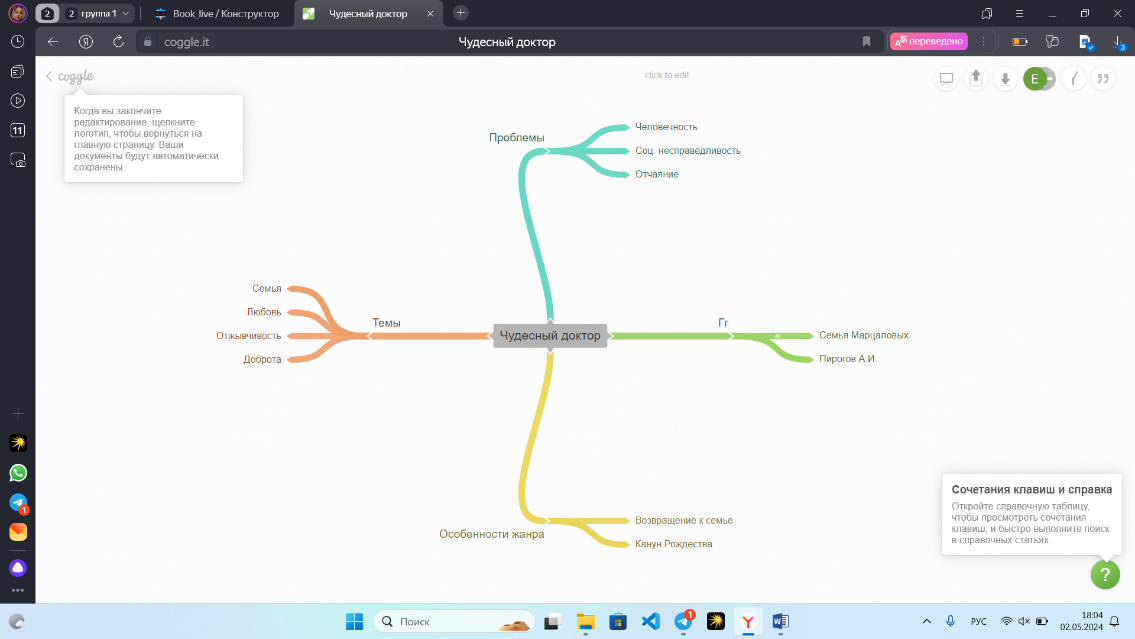
Чтобы построить граф в программе Easy Linavis я взяла произведение Толстого «Власть тьмы» и выделила основных персонажей из каждой главы.

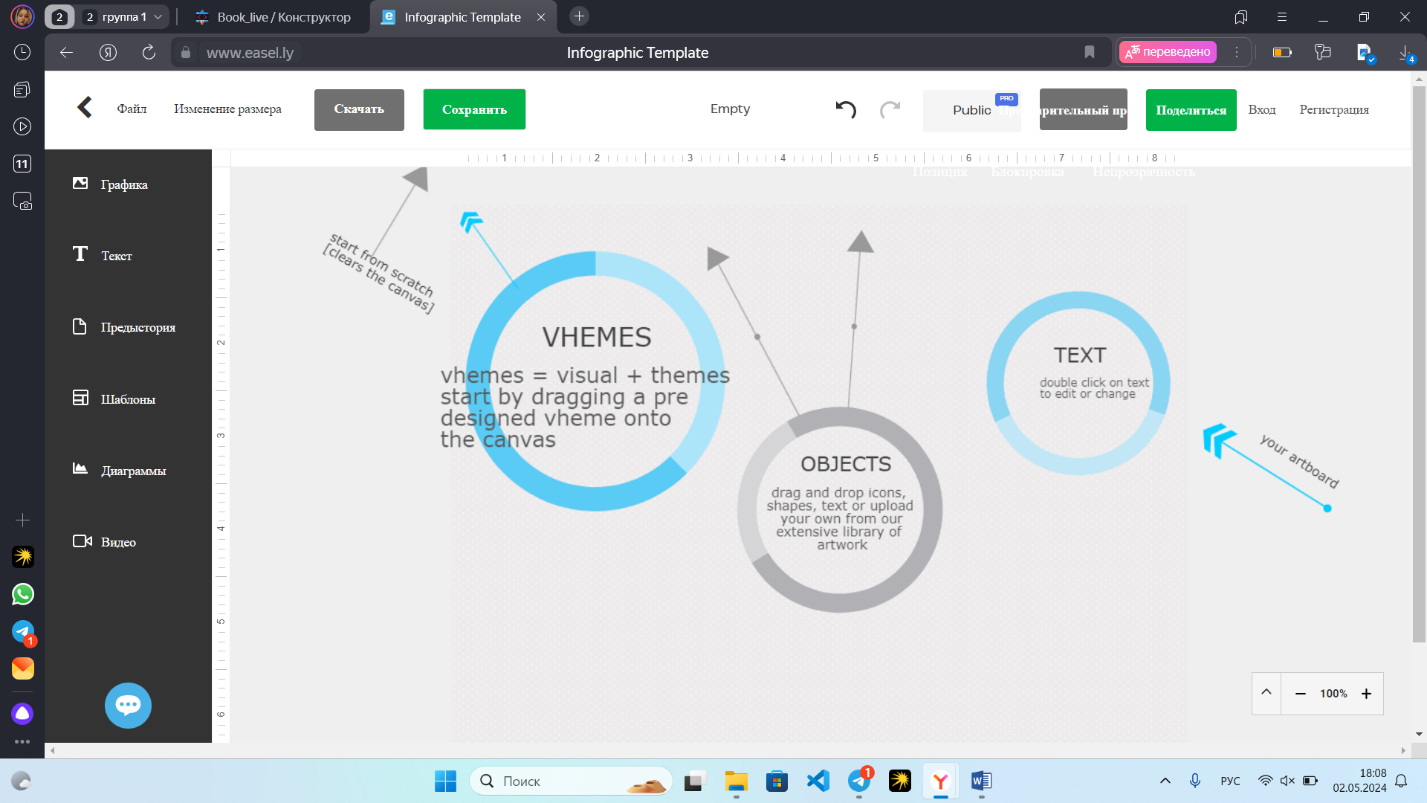


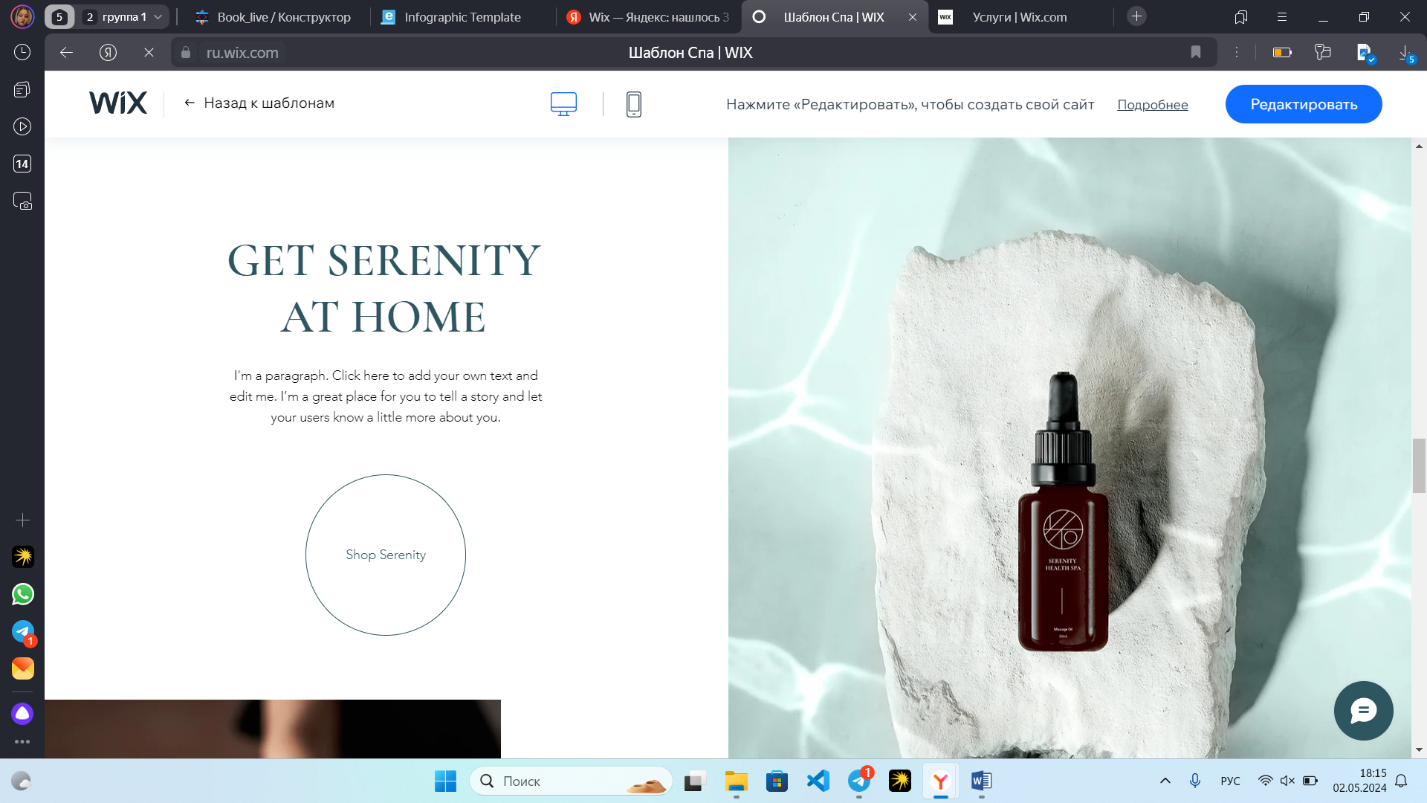
**Уровень 2.**

**Задание 2.**

**1.** Чтобы проверит ресурс Coogle я взяла произведение Куприна «Чудесный доктор».



**2.** В ресурсе easel.ly можно разработать с помощью шаблонов свой пост, построить диаграммы.**3.** С помощью ресурса Wix можно создать свой сайт используя шаблоны или самому.



**Задание 3.**

1. Ранг: Ранг - это понятие, используемое для определения порядка частотности слов или языковых единиц в тексте. Чем выше ранг, тем более часто встречается данное слово. Ранг можно определить как порядковый номер слова в списке по убыванию частотности.

2. Относительные частоты: Относительные частоты представляют собой процентное соотношение частоты конкретного языкового элемента к общему числу языковых элементов в тексте. Они помогают оценить важность или распространенность конкретного элемента в тексте.

3. Показатель R (range): Показатель R (range) используется для оценки разнообразия или разнообразия лексики в тексте. Он вычисляется как отношение числа уникальных слов к общему числу слов в тексте.

4. Коэффициент Жуйяна: Коэффициент Жуйяна - это метрика, которая оценивает степень равномерности распределения частотности слов или языковых единиц в тексте. Он рассчитывается на основе графика кумулятивной частоты слов.

5. ARF (average relative frequency): ARF - это средняя относительная частота, которая показывает среднюю частотность конкретного языкового элемента в тексте. Она вычисляется как сумма относительных частот всех элементов, деленная на их количество.

6. Критерий логарифмического правдоподобия: Критерий логарифмического правдоподобия используется для оценки степени соответствия наблюдаемых данных распределению вероятностей. Он позволяет определить, насколько хорошо модель соответствует данным.

7. Критерий Йенсена-Шеннона: Критерий Йенсена-Шеннона используется для измерения информационной энтропии в тексте или корпусе. Он позволяет оценить степень неопределенности или разнообразия языковых данных.

**Задание 4.**

**1.** Формула Сводеша выглядит следующим образом:

t = 1/2log\_2 ( D/d)

где:

- t - время разделения двух языков в годах,

- D - количество лексических элементов (слов) в общем словаре двух языков,

- d - количество общих лексических элементов (слов) в словаре двух языков.

Формула Сводеша позволяет оценить время разделения двух языков на основе количества общих и уникальных слов в их словарях. Однако следует помнить, что данная формула имеет свои ограничения и не всегда точно отражает историческую реальность разделения языков.

**2.** Основная идея формулы Старостина заключается в том, что чем больше общих слов у двух языков, тем более близкими они считаются и тем меньше времени прошло с момента их разделения.

Формула Старостина выглядит следующим образом:

t = - log N\_12/log N\_0

где:

- t - время разделения двух языков в годах,

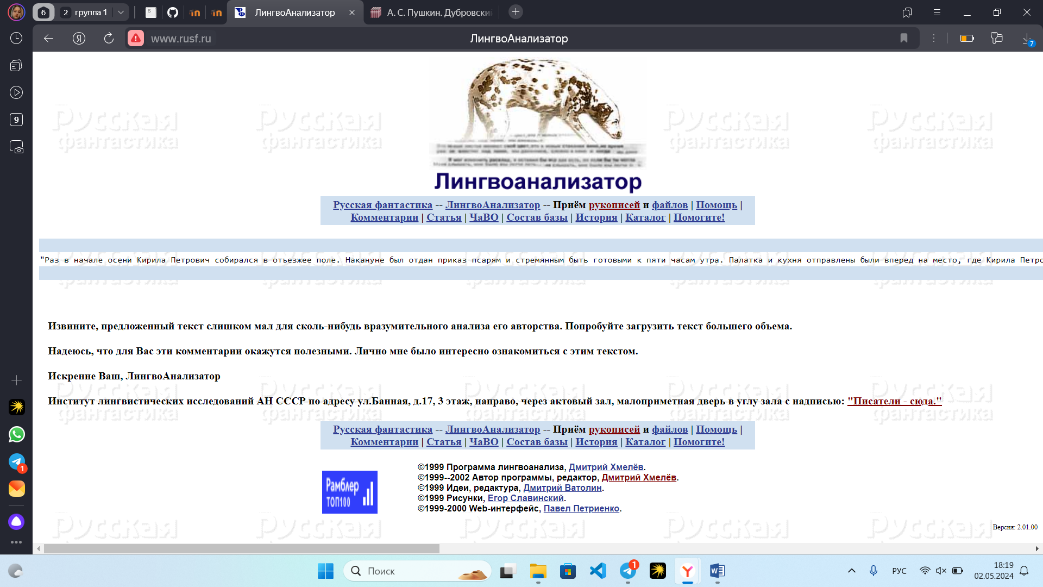
- N\_12 - количество общих лексических элементов (слов) в словаре двух языков,

- N\_0 - общее количество лексических элементов (слов) в словаре двух языков.

Формула Старостина также позволяет оценить время разделения двух языков на основе количества общих слов в их словарях. Однако как и другие методы глоттохронологии, она имеет свои ограничения и не всегда точно отражает историческую реальность разделения языков.

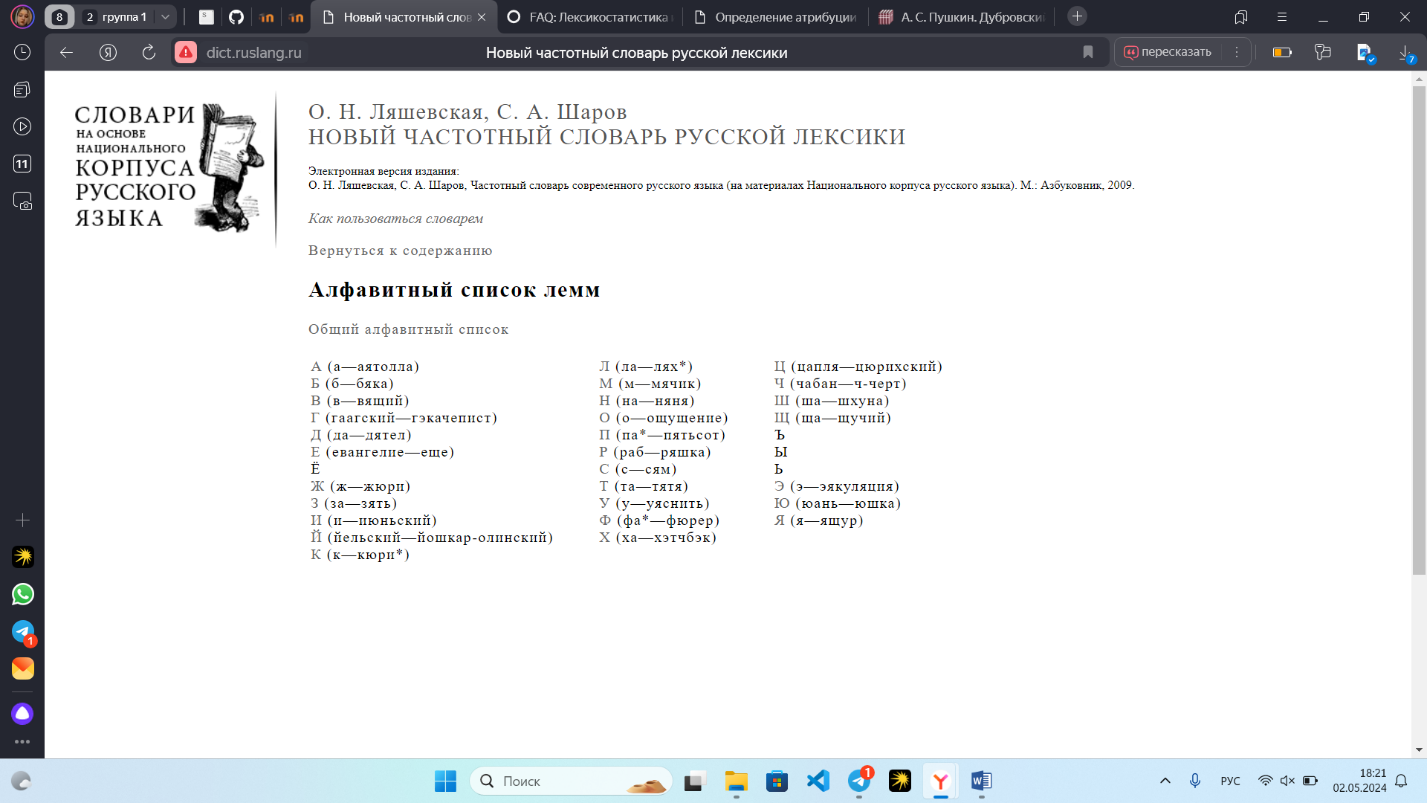
**Задание 5.**

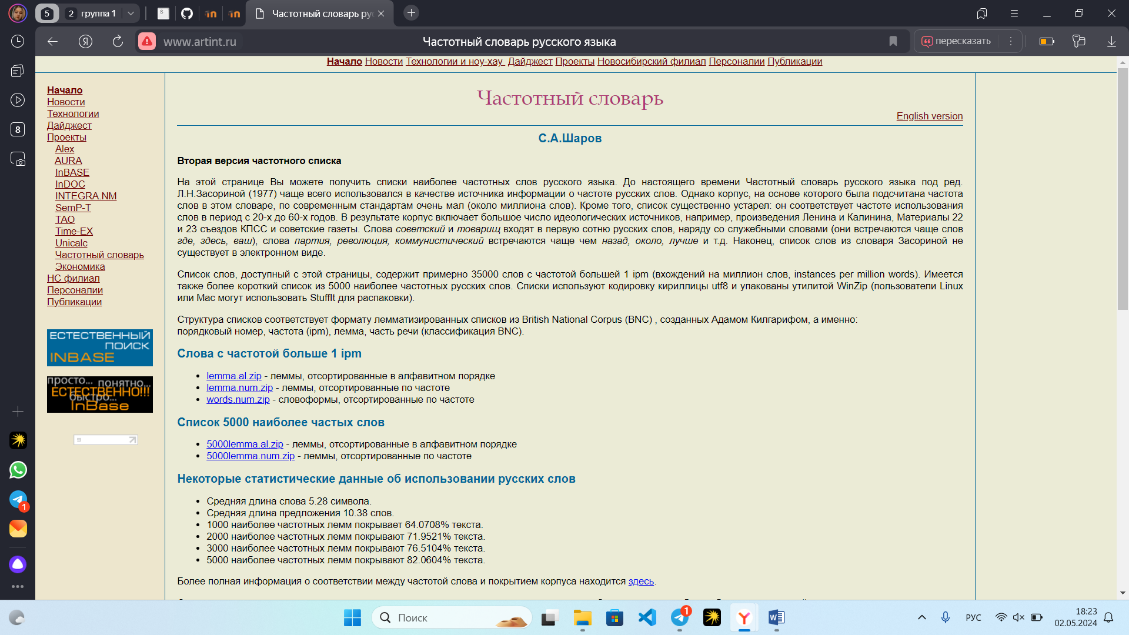
**1.** Для проверки сайта ЛингвоАнализатор я взяла произведение Пушкина «Дубровский»



Сайт не справился с задачей.

**2.**

**4.**



**5.**

