ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

Воронежский государственный университет

Факультет романо-германской филологии

Кафедра испанской филологии

Направление «Лингвистика»

Профиль «Теория и методика преподавания иностранных языков и культур»

Проект на тему: Визуализация данных и сетевой анализ: облака тегов, Voyant Tools, Gephi. Анализ социальных сетей (построение графов)

Выполнила: студентка 1 курса

Гришанкова Виктория

Преподаватель: Донина О.В.

Воронеж, 2018 год.

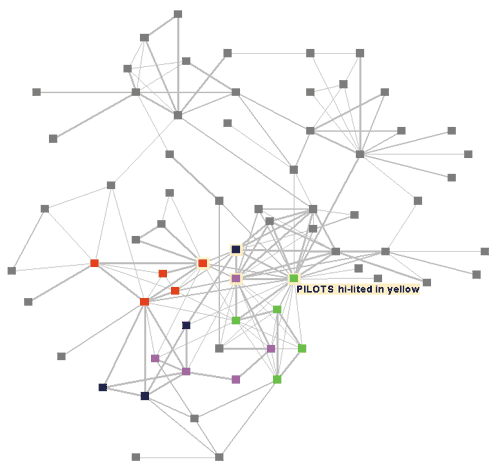
**Визуализация данных** — это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению Визуализация данных находит широкое применение в научных и [статистических](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0) исследованиях (в частности, в прогнозировании, [интеллектуальном анализе данных](https://ru.wikipedia.org/wiki/Data_mining), [бизнес-анализе](https://ru.wikipedia.org/wiki/%D0%91%D0%B8%D0%B7%D0%BD%D0%B5%D1%81-%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7)), в [педагогическом дизайне](https://ru.wikipedia.org/w/index.php?title=%D0%9F%D0%B5%D0%B4%D0%B0%D0%B3%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B4%D0%B8%D0%B7%D0%B0%D0%B9%D0%BD&action=edit&redlink=1) для обучения и тестирования, в новостных сводках и аналитических обзорах. Визуализация данных связана с [визуализацией информации](https://ru.wikipedia.org/w/index.php?title=%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8&action=edit&redlink=1), [инфографикой](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%84%D0%BE%D0%B3%D1%80%D0%B0%D1%84%D0%B8%D0%BA%D0%B0), [визуализацией научных данных](https://ru.wikipedia.org/w/index.php?title=%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%D0%BD%D0%B0%D1%83%D1%87%D0%BD%D1%8B%D1%85_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85&action=edit&redlink=1), [разведочным анализом данных](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D0%B7%D0%B2%D0%B5%D0%B4%D0%BE%D1%87%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85) и [статистической графикой](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%B3%D1%80%D0%B0%D1%84%D0%B8%D0%BA%D0%B0).

# Визуализация социального графа

В математическом смысле социальная сеть — не что иное как граф с вершинами и ребрами, которые, соответственно, представляют акторов (субъектов общения) и связи между ними.

**граф** — это совокупность объектов со связями между ними. Объекты представляются как вершины, или узлы **графа**, а связи — как дуги, или рёбра.

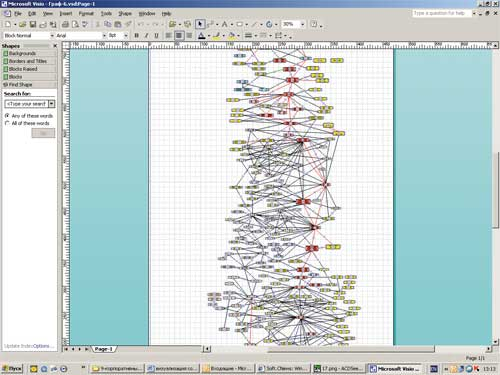
Вашему вниманию — ставший уже классическим пример с построением графа коммуникаций между пилотами самолетов 11 сентября 2001 года на основании фактов переговоров по линиям радиосвязи:



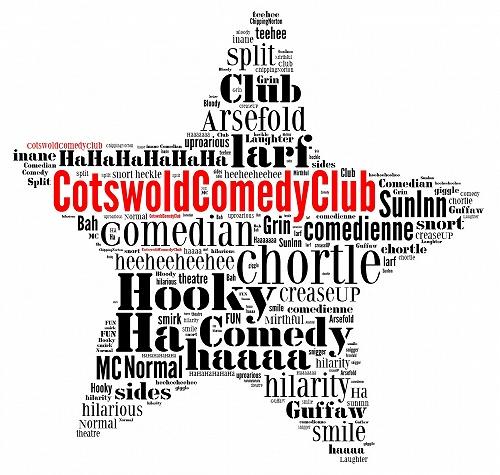
Каждая вершина графа обозначала отдельного пилота, каждое ребро, соединяющее две вершины, – факт переговоров между пилотами. Очевидно, что вывод о том, какие пилоты являлись террористами, делался на основании наблюдения, что в некоторых местах графа плотность связей (количество переговоров) возрастало.

# Инструментарий визуализации графов

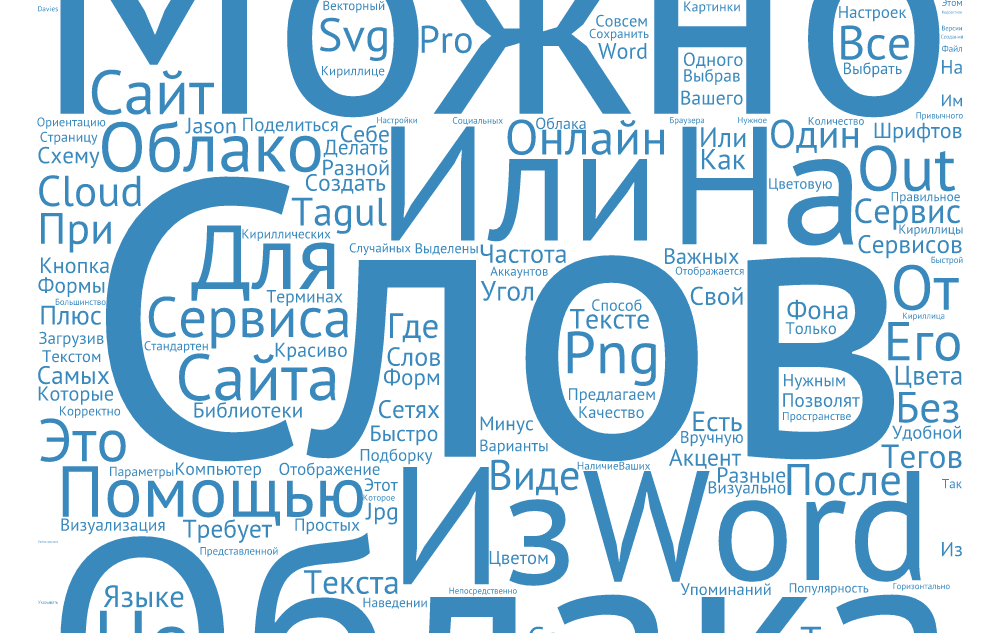
В общем случае, инструменты визуализации графов делятся на две категории – это программы, позволяющие строить графы вручную (специализированные редакторы) и программы, оснащенные инструментарием и интерфейсом для автоматизированной визуализации графа на основании введенных данных. Конечно же, существует и третий вариант – многие современные языки и среды программирования (например, Java с библиотекой JUNG) имеют в распоряжении библиотеки, пользуясь которыми, можно создать пользовательский интерфейс, отвечающий определенным требованиям. Но сегодня мы остановимся на уже готовом инструментарии, который на данный момент доступен специалистам (и неспециалистам) по большим данным.

Microsoft Office Visio – один из наиболее популярных инструментов ручного построения разнообразных схем и диаграмм – в том числе, графов. Средство обладает всеми нужными широкому кругу пользователей функциями, относительно удобно в использовании 

**облако тегов (tag cloud)** - представляет собой множество ключевых слов или словосочетаний – тегов, извлеченных из текста, изображенных на плоскости. Размер каждого тега зависит от частоты или любой другой частотной характеристики тега. Облако тегов может иметь любую форму: действительно облака или, например, звездочки.



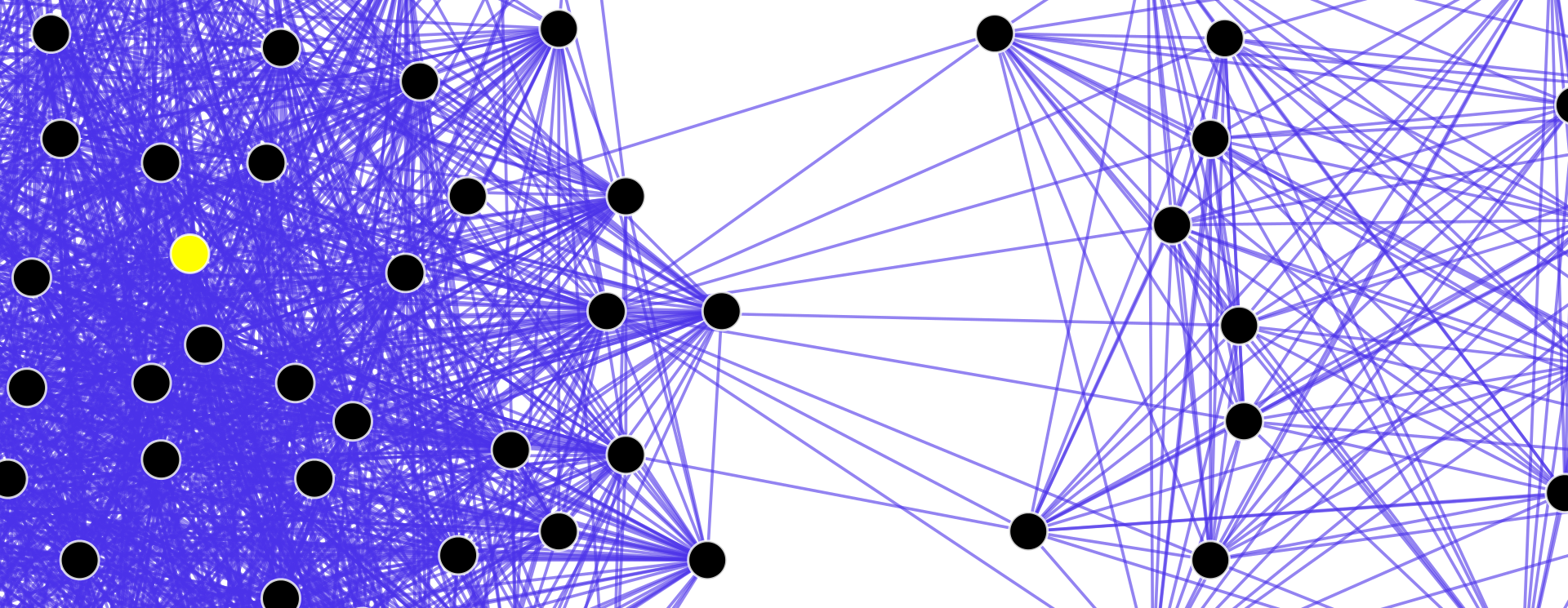
Частота слов в статье, акцент на важных терминах или популярность упоминаний в тексте – все это можно сделать с помощью облака слов



# СЕТЕВОЙ АНАЛИЗ

# **Анализ социальных сетей** (*Social network analysis или SNA*) представляет собой исследование социальных сетей, рассматривающее социальные отношения в терминах теории сетей. Эти термины включают в себя понятие *узла* (отображает отдельного участника в пределах сети) и *связи* (отображает такие отношения между индивидами, как дружба, родство, положение в организации, интимные отношения, и т. д.). Эти сети часто описывают в виде социальных сетевых схем, где узлы представлены в виде точек, а связи представлены в виде линий

В основе анализа социальных сетей лежит математическая теория графов (она представлена в работах таких авторов, как Эрдос, Харари и Раппапорт), а также эмпирические исследования в области социальной психологии и антропологии (Хайдер и Морено). В то время как первая группа учёных открывала различные законы построения абстрактных узлов и линий, последние обнаружили, что узлами и линиями удобнее всего обозначать отношения между людьми. Поскольку обе группы исследователей работали в одно и то же время (во второй половине двадцатого века), они сошлись на разработке серии метрик и методов, позволяющих выделять базисные структуры сложных эмпирических явлений.



# 

# 5 онлайн-сервисов для создания облака тегов и слов на русском языке

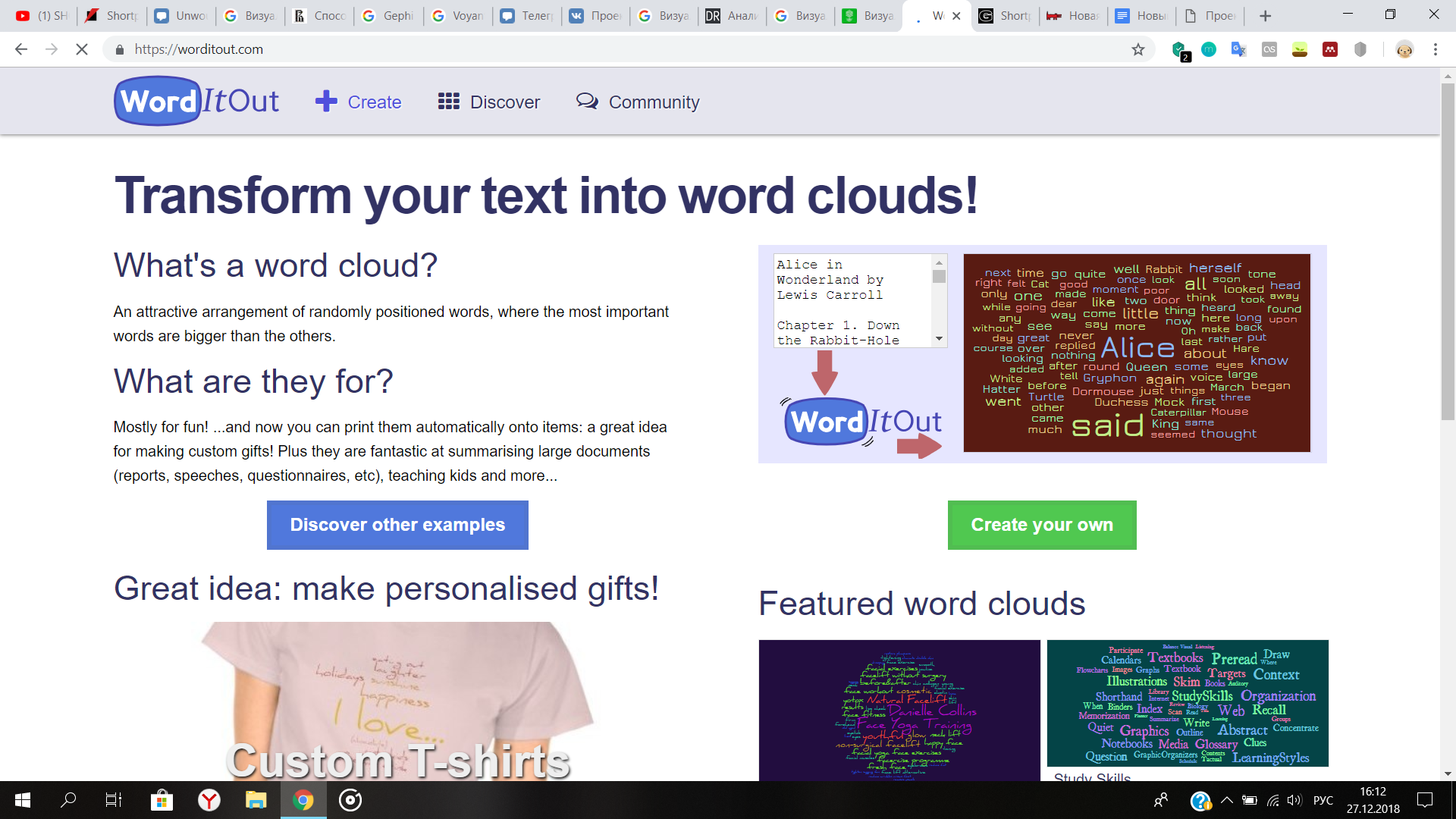
### 1. Word It Out

### Это один из самых простых сервисов, который позволяет делать облако слов без регистрации из вводимого текста или ссылки.

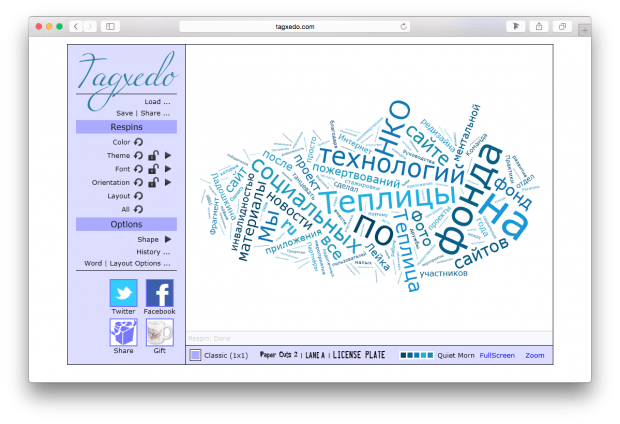
В настройках можно поменять шрифт, цвета слов и фона. Удобной является кнопка случайных настроек, с помощью которой можно генерировать разные варианты. Плюс сервиса – правильное отображение кириллических шрифтов.

После создания облака его можно скачать в виде изображения или поделиться с помощью ссылки.

Сайт: [Word It Out](http://worditout.com/word-cloud/make-a-new-one).



### 2. Tagxedo



ФРАГМЕНТ САЙТА TAGXEDO.

### Сервис не требует регистрации и позволяет создавать облака слов разной формы из представленной библиотеки, загрузив текст или указав ссылку на страницу с нужным текстом.

В настройках можно изменить цветовую схему, параметры текста, а также ориентацию слов в пространстве. Кириллица отображается корректно.

После создания облака слов его можно сохранить к себе на компьютер в виде картинки в формате jpg или png, выбрав нужное качество.

Сайт: [Tagxedo](http://www.tagxedo.com/app.html).

### 3. TagulФрагмент сайта Tagul.

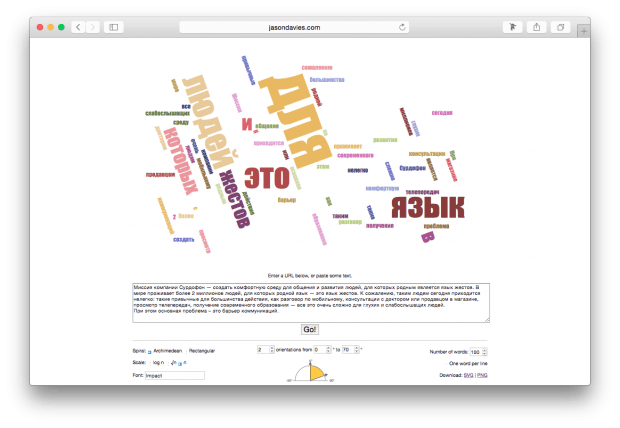
ФРАГМЕНТ САЙТА TAGUL.

### Этот сервис также позволяет создать облако слов непосредственно в окне вашего браузера, но требует быстрой регистрации с помощью одного из ваших аккаунтов в социальных сетях.

Текст также можно добавить вручную или указать ссылку. Настройки позволяют изменить большинство параметров, а наличие библиотеки форм – создать облако слов, которое будет отличаться от привычного. Для правильного отображения кириллицы необходимо выбрать один из поддерживаемых шрифтов (выделены синим цветом).

После завершения создания облако слов можно скачать на свой компьютер в формате png, а также поделиться им с помощью [ссылки](https://tagul.com/show/tprix3c3aqre/new)

4. Word Cloud

ФРАГМЕНТ САЙТА WORD CLOUD.

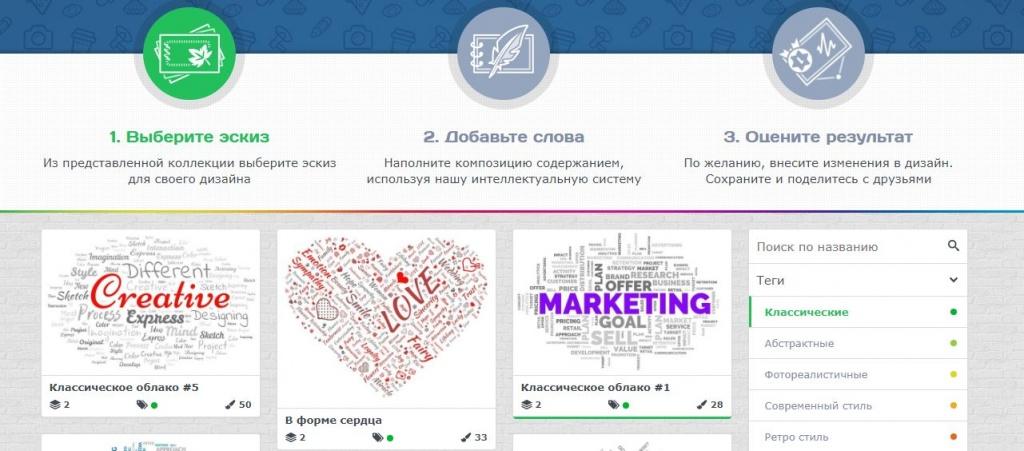
### Онлайн-сервис от Jason Davies для создания облака слов, где можно указывать угол расположения слов, а не только вертикально и горизонтально.

Также можно поменять шрифт, указать количество слов для составления облака, а результат можно скачать как в формате png, так и в svg.

Минус этого сервиса – не совсем корректное отображение слов на кириллице. При этом способ загрузки слов стандартен: можно ввести текст или указать ссылку.

Сайт: [Word Cloud](https://www.jasondavies.com/wordcloud/).

### 5. Wordcloud.pro

ИЗОБРАЖЕНИЕ: СКРИНШОТ С САЙТА WORDCLOUD.PRO/RU

### С помощью сервиса можно сделать облако слов, открытку в виде поздравления, рекламный баннер или постер.

Всего нужно выполнить несколько шагов: зарегистрироваться на сайте сервиса, выбрать понравившийся шаблон, добавить нужные слова (можно воспользоваться доступной коллекцией текста) и сохранить дизайн для печати или для социальных сетей. Есть базовый, бесплатный тариф. В нем доступна коллекция эскизов, можно использовать для личных и коммерческих целей. Также вы может [прочитать подробную инструкцию](http://wordcloud.pro/ru/blog/step-by-step-guide-for-creating-word-clouds) в блоге студии о создании облака слов.

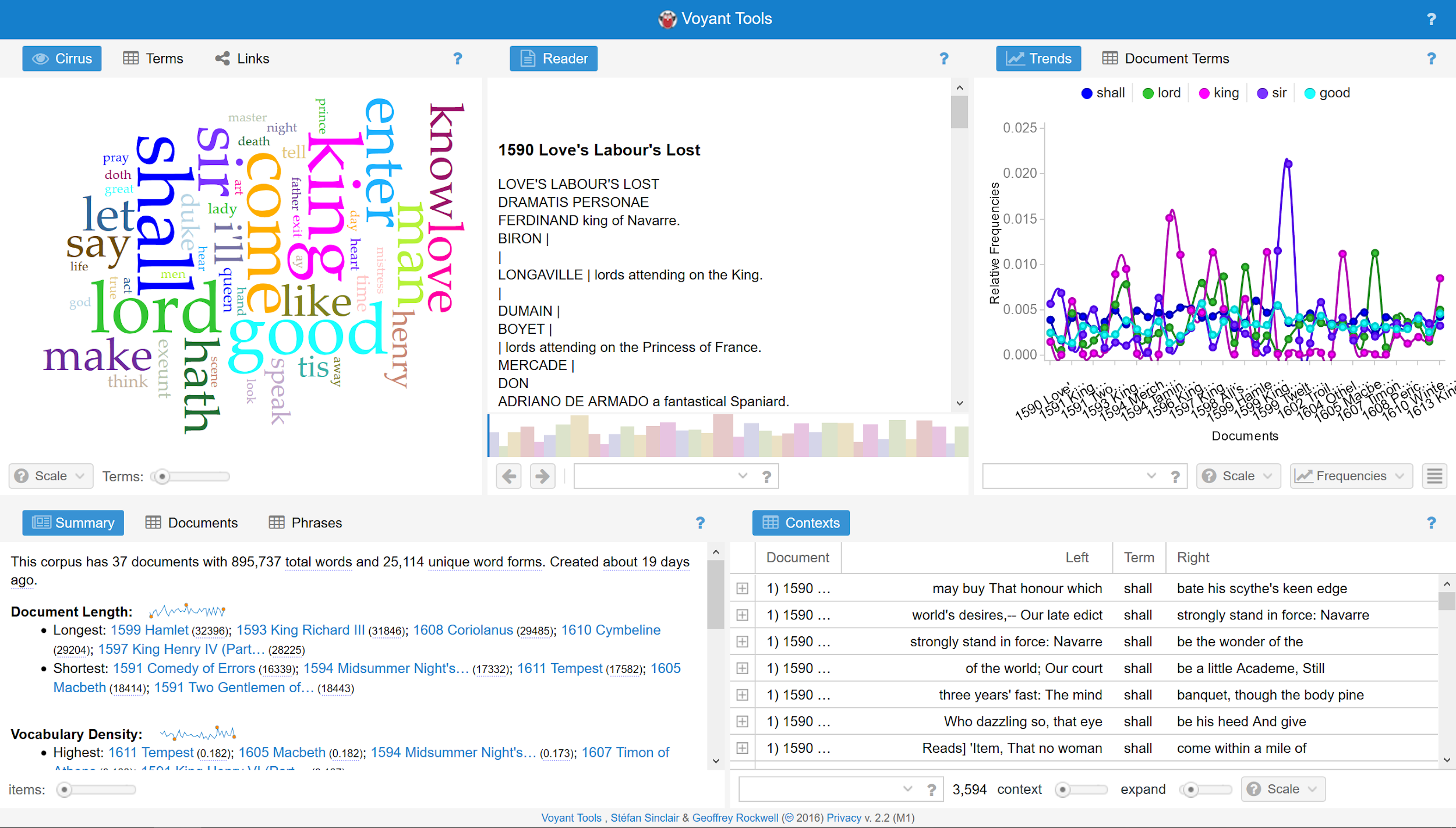
6. Voyant Tools

Voyant Tools - веб-приложение для анализа текстов

Инструкцию по работе с ресурсом вы можете найти, перейдя по [ссылке](http://docs.voyant-tools.org/start/)

Что позволяет платформа *Voyant tools* и для чего она может быть использована:

* изучить возможности компьютеризованного анализа. Инструменты, обеспеченные платформой, помогают выполнять многие академические задачи
* изучить и проанализировать тексты, существующие как в веб-пространстве, так и на компьютере пользователя;
  + добавить функциональности онлайн-коллекциям, каталогам, цифровым архивам, блогам, сайтам: позволяет другим увидеть пользовательские тексты через призму аналитических инструментов;
  + сформировать интерактивный графический материал для онлайн-публикуемых эссе. Позволяет дополнить интерактивными вставками те исследования, которые публикуются в веб для того, чтобы читатель смог легко проанализировать публикуемые материалы и даже самостоятельно воспроизвести их результаты;
  + разработать свой набор инструментов на базе существующего кода.



7. Gephi

Ресурс используется для построения графов

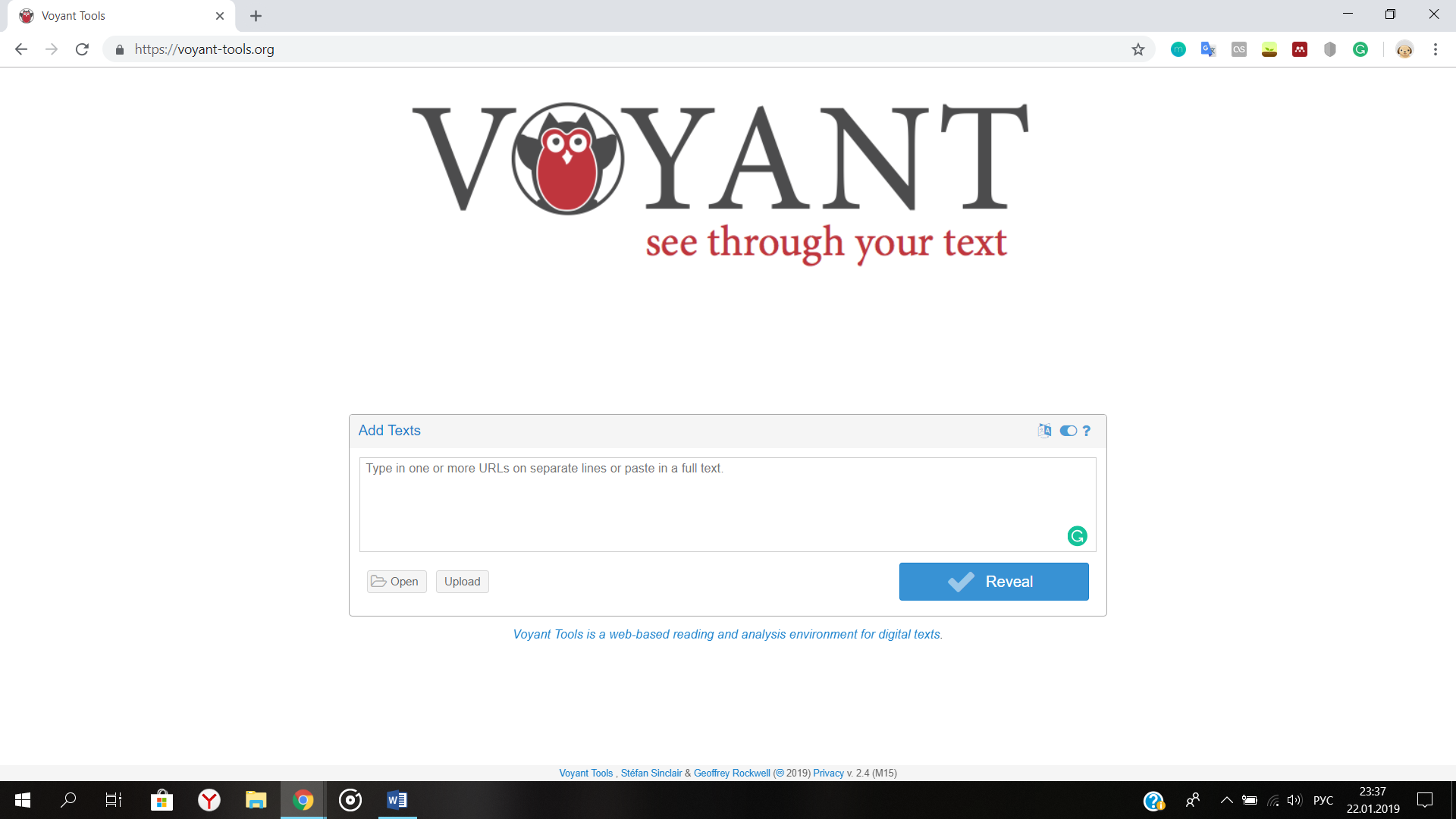


# 

# Работа с word it out

1. Откройте главную страницу <https://worditout.com/word-cloud/create> ресурса
2. Откройте любую понравившуюся вам статью. Мне понравилась [эта](https://arzamas.academy/mag/616-gogols_mistakes)
3. Скопируйте документ и вставьте в поле Original text. Нажмите Generate. Получится что-то подобное:

# Работа с Voyant Tools

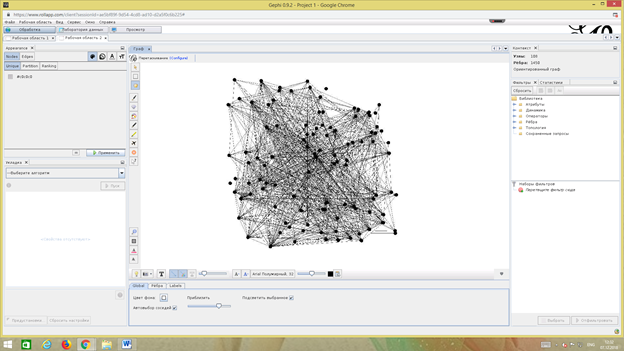
Проделаем то же самое толькос voyant tools. Вставьте текст в специальное окно и нажмите reveal 

# Работа с Gephi

Скачиваем gephi по [ссылке](https://gephi.org/users/download/)

Постройте граф своих друзей, используя [приложение](https://vk.com/app3861133)

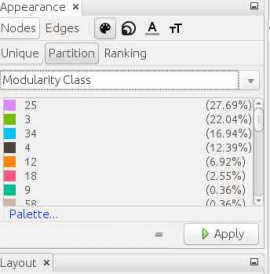
Импортируем выгруженный список друзей. Граф должен выглядеть примерно вот так:



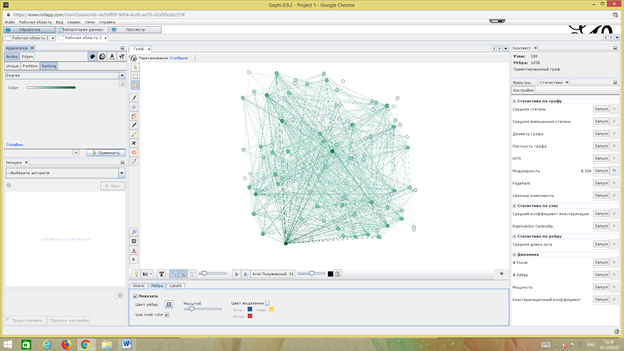
Управляем масштабом графа при помощи колеса вашей мыши, а так же при помощи значка лупы внизу

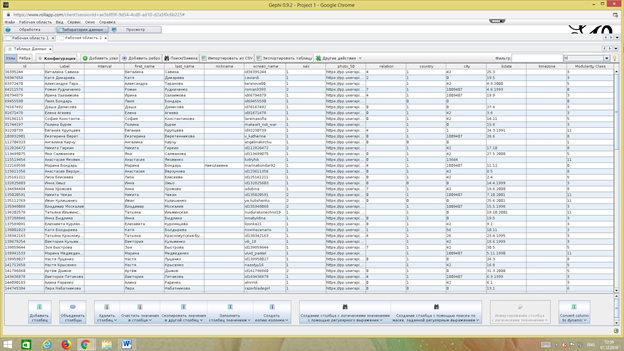
После импорта для наглядности Измените внешний вид графа: запустите алгоритм укладки (раздел "Window"->"Layout"), подберите оптимальные значения параметров, отредактируйте ширину и цвет ребер графа, цвет и размер узлов.

Рассчитаем модулярность графа (раздел "Window"->"Statistics"->"Modularity") и раскрасьте граф по классам модулярности (раздел "Window"->"Appearance"->"Nodes"->"Partition").

Во вкладке "Preview" выберите фон, цвета подписей, размер узлов и.т.д

Поскольку узлы и ребра раскрашены одним цветом (Unique), мы раскрасим их в соответствии с тем, к каким сообществам они принадлежат (Partition > Modularity Class).

Далее делаем укладку ваших данных. Меню "Укладка" находится в левом нижнем углу. Вы можете прервать ее в любой момент, т.к. укладка может идти вечно. 

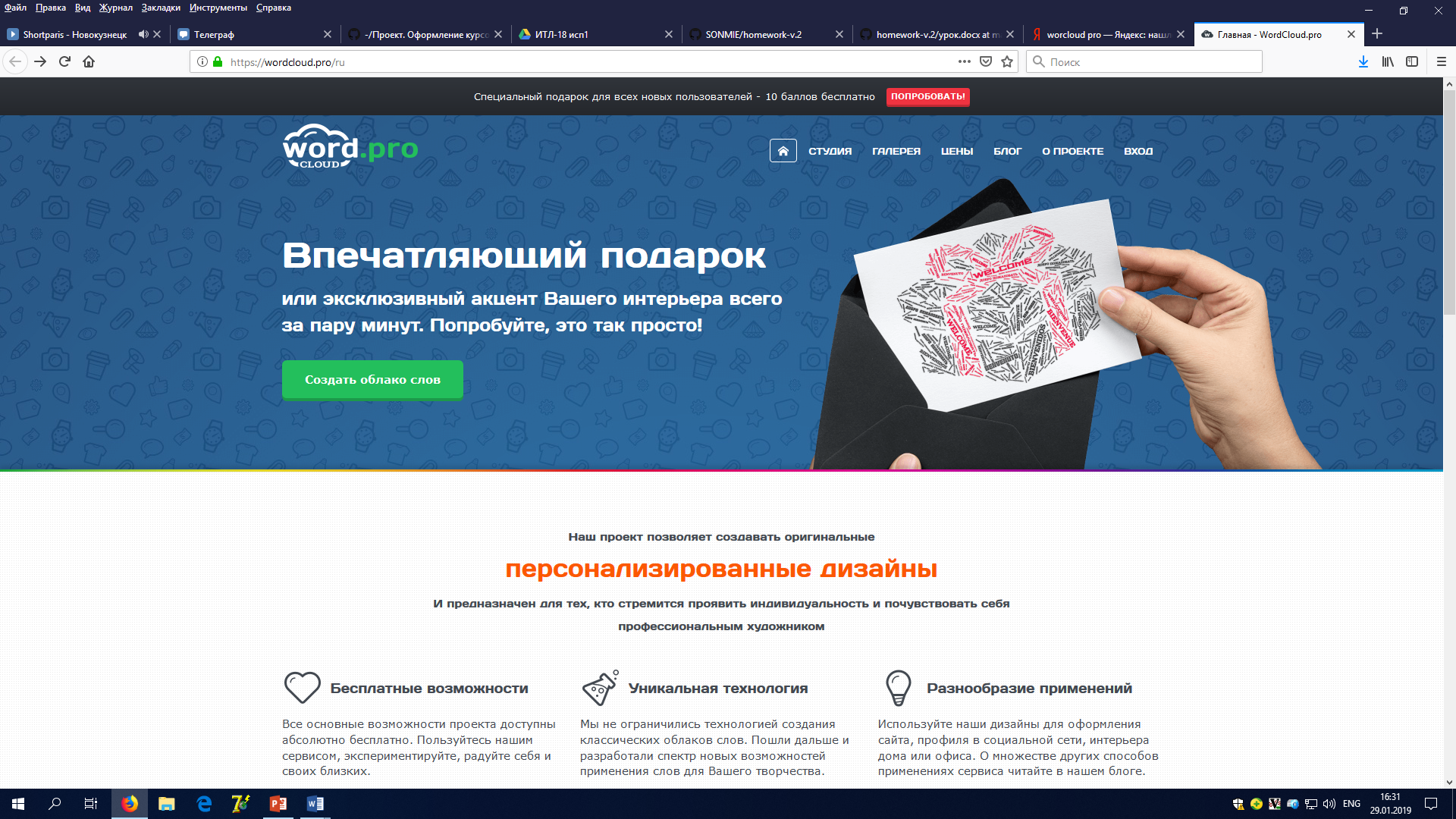
Переходим во вкладку Лаборатория данных, которая позволяет рассмотреть ваши данные в таблице. 

Сохраните граф в формате png или в pdf

**Работа с WORDCLOUD.PRO**

Зайдите на wordcloud.pro и создайте облако слов.

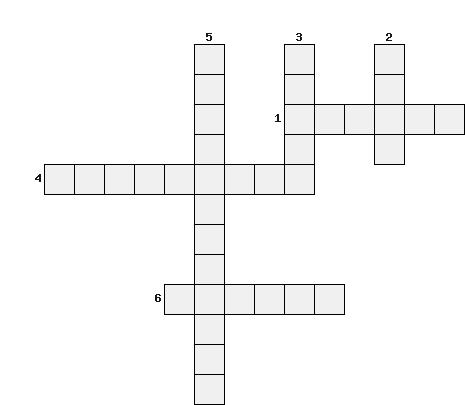
Учтите, что перед этим действием нужно зарегистироваться на сайте по адресу электронной почты!



# 

# Контрольные вопросы

1. Что такое облако тегов?
2. Что такое визуализация данных?
3. С помощью каких онлайн-ресурсов можно построить облако тегов?
4. С помощью каких онлайн-ресурсов можно построить граф?
5. Почему визуализация данных удобна?

**Кроссворд для закрепления теории:**

**Задания для кроссворда:**

1. Представляет собой множество ключевых слов или словосочетаний – тегов
2. Это совокупность объектов со связями между ними
3. Из чего состоит облако?
4. Инструмент визуализации графов
5. Это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению

6. То, что представляется в визуализации.

Ответы: 1- облако; 2-граф; 3-слова; 4-программа; 5-визуализация; 6-данные

Заполните таблицу ресурсами для построения графов

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**Соотнесите:**

**ОБЛАКО ТЕГОВ** это совокупность объектов со связями между ними

**ГРАФ**

это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению

**ВИЗУАЛИЗАЦИЯ ДАННЫХ**

представляет собой множество ключевых слов, извлеченных из текста, изображенных на плоскости

**Продолжите определение:**

**облако тегов -**

**граф-**

**визуализация данных-**

**с помощью каких ресурсов можно образовать облако тегов? Укажите не менее 3:**

**1)**

**2)**

**3)**

**Использованные ресурсы:**

[**https://te-st.ru/2015/05/13/make-word-cloud-online/**](https://te-st.ru/2015/05/13/make-word-cloud-online/)

[**http://datareview.info/article/analiz-sotsialnyih-setey-instrumentyi-vizualizatsii-grafa/**](http://datareview.info/article/analiz-sotsialnyih-setey-instrumentyi-vizualizatsii-grafa/)

[**https://ru.wikipedia.org/wiki/Визуализация\_данных**](https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85)