

Цифровая грамотность

семинар 10

Voyant tools

[Voyant Tools](#) -- веб-приложение для анализа текста, который можно загрузить из файла, нескольких файлов или по ссылке. Оно поддерживает много языков и позволяет рассчитать частотность слов, построить по ним облака тегов и красивые графики.

Voyant tools

Примеры исследований, сделанных с помощью Voyant tools:

1. <http://www.maxkemman.nl/2016/11/a-republic-of-emails-what-are-the-contents/>
2. <http://britishlibrary.typepad.co.uk/digital-scholarship/2013/05/on-metadata-and-cartoons.html>

Больше здесь:

<http://docs.voyant-tools.org/about/examples-gallery/>

Voyant tools

1. Перейдите по ссылке: <https://voyant-tools.org>
2. Загрузите документ “капитанская дочка”
3. Отредактируйте список стоп-слов. Возьмите яндексовский список стоп-слов
rus_stepwords.txt

Voyant tools

Посмотрите различные функции:

1. Cirrus - инфографика с наиболее частотными словами
2. Terms - список слов
3. Collocations - анализ контекста
4. Links - граф контекста
5. Phrases - повторяющиеся фразы
6. Trends/Bubbleline - наиболее частые слова в отдельных сегментах документа

Voyant tools

Выберете произвольный текст, добавьте стоп-слова, постройте соответствующую визуализацию и ответьте на следующие вопросы:

1. Какие слова встречаются чаще?
2. Какие наиболее частые 2, 3 - граммы?
3. Проанализируйте трендовую линию по наиболее частым словам в сегментах документа
4. Постройте несколько различных визуализаций, предлагаемых сервисом. Например, Knots :)

Еще немного про облака тегов

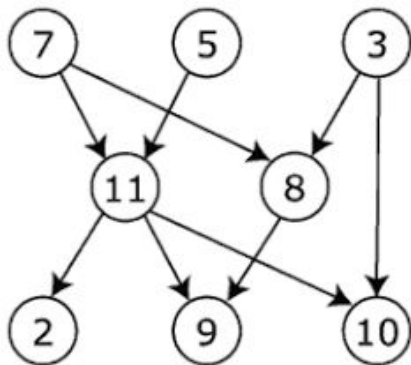
Облако тегов (*word cloud*) -- это визуализация ключевых слов (тегов) из какого-либо текста, где размер слова указывает на его частотность. Такая картинка украсит любую научную работу. :)

Еще немного про облака тегов

Существует много простых и удобных онлайн-сервисов для создания облаков тегов; вот некоторые из них.

- [Word Clouds](#) - тонкие настройки, разные формы облаков, фильтрация стоп-слов, можно загрузить файл
- [Wordle](#) - настройка формы облака, шрифта и цвета слов, работа с небольшими текстами
- [Word It Out](#) - настройки цвета и размера слов, работа с небольшими текстами
- [Tagxedo](#) - настройка ориентации слов, цветовой схемы, формы облака
- [Word Cloud by Jason Davies](#) - настройка количества слов и их ориентации

Граф



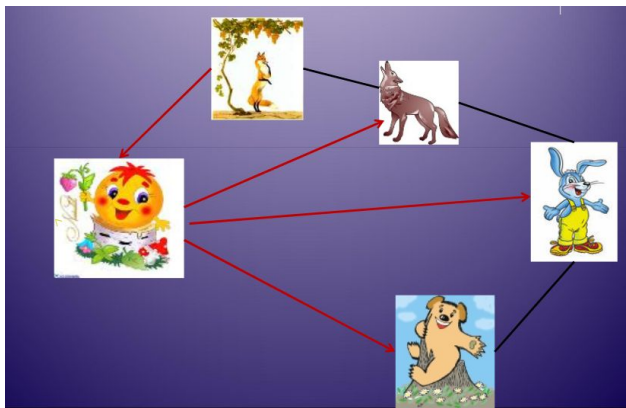
Граф - это множество точек или вершин и множество линий или ребер, соединяющих между собой все или часть этих точек. Вершины, прилегающие к одному и тому же ребру, называются смежными.

Если ребра ориентированы, что обычно показывают стрелками, то они называются дугами, и граф с такими ребрами называется ориентированным графом. Если ребра не имеют ориентации, граф называется неориентированным.

Зачем нужны?

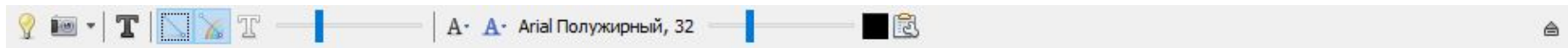
Очень хорошо подходят для анализа данных с сложными взаимосвязями.

1. граф-схемы алгоритмов
2. В коммуникационных и транспортных системах
3. для описания структур, путей сложных реакций в химии
4. и.т.д



Gephi - инструмент для визуализации графов

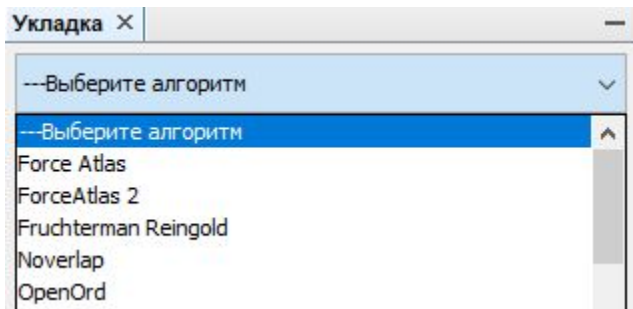
- 1) Установите Gephi (<https://gephi.org/>)
- 2) Скачайте и откройте lesmiserable.gexf - взаимосвязи героев романа Виктора Гюго “Отверженные”.
- 3) Управление: правая кнопка мыши для перемещения, средняя - масштаб.
- 4) На нижней панели можно отрегулировать толщину ребер и подписей



толщина ребер

Алгоритмы Force Atlas, Fruchterman Reingold

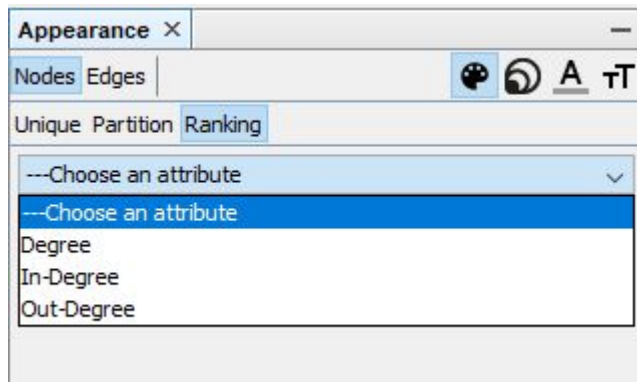
Идеи этих алгоритмов - в наложении сил на ребра и вершины графа, чтобы минимизировать число пересекающихся ребер, а также выровнять ребра по длине.



Запустите алгоритм укладки, попробуйте эти алгоритмы и разные параметры. Попробуйте поменять силы отталкивания и притяжения. Как они меняют вид графа? Какой набор параметров лучше всего?

Ранги вершин

Степенью вершины называется число ребер графа, которым принадлежит эта вершина. Раскрасьте граф по степеням вершин.



Статистики

Модулярность — количественный параметр использующий количество внутренних связей внутри предполагаемых сообществ и связей с внешними сообществами. Рассчитайте модулярность графа и раскрасьте, используя этот параметр.

Компонента связности графа — максимальный связный подграф. Рассчитайте компоненты связности и раскрасьте, используя этот параметр.

Какие выводы можно сделать, используя эти параметры?

Фильтры	Статистики X	
Настройки		
▼ Статистика по графу		
Средняя степень	Запуск	●
Средняя взвешенная степень	Запуск	●
Диаметр графа	Запуск	●
Плотность графа	Запуск	●
НITS	Запуск	●
Модулярность	Запуск	●
PageRank	Запуск	●
Связные компоненты	Запуск	●

Предпросмотр и экспорт

Поменяйте на свой вкус и цвет фон/размеры узлов/ребер/раскраску графа.

Попробуйте различные способы экспорта графа из gerhi.

PDF/SVG/PNG

Спасибо за внимание!