1) ЦЕЛЬ И ЗАДАЧИ

Цель:
Сделать сайт https://rusvectores.org более удобным и информативным для исследователей-лингвистов и general public.

Задачи:
Исследовательские:

1. Найти более удобное отображение лингвистической информации, извлечённой из векторных моделей языка.
2. Сравнить данные семантической разметки НКРЯ и близость слов в моделях.

Инженерные:
Необходимо реализовать:

1. фильтры по частотности
2. фильтры по частям речи
3. фильтры на другие метаданные по семантическим классам
4. старый список ближайших соседей → граф с делением на кластеры, с указанием на близость слова к центру кластера. Графы должны быть динамическими, с картинками и ссылками, с возможностью перетаскивать узлы, кликнуть на узел и получить граф слова
5. Опционально: большая семантическая карта
6. Опционально: задавать семантический класс из НКРЯ [Надя: представляю это как выдачу визуализации на запрос вида “НКРЯ -> части растений (в окошке с прокруткой и поиском по списку)”]

2) ДАННЫЕ И МЕТОДЫ

Векторные модели корпусов:

1. Корпусы:
    a. НКРЯ
    b. Википедии
    c. Поток новостей с 1 500 преимущественно русскоязычных новостных сайтов (около 30 миллионов документов);
    d. Araneum Russicum Maximum
    e. Тайга
    f. Веб: случайно отобранные 9 миллионов русскоязычных веб-страниц, скачанных в декабре 2014 года.
2. Семантические метки НКРЯ
3. Возможно, семантическая разметка из проекта: https://research.kartaslov.ru/
    a. Открытая семантика русского языка — разметка слов и выражений русского языка по семантическим срезам («люди», «животные», «сооружения», «вещи», «действия» и т.д.).

b. Тональный словарь русского языка — слова и выражения русского языка, размеченные по полярности (положительная, отрицательная, нейтральная). Также приводится сила выраженности эмоционально-оценочного заряда.
c. Ассоциации к словам и выражениям русского языка — ассоциации к словам и выражениям русского языка, придуманные реальными людьми. Кроме общего набора публикуются данные срезов по гендеру, т.е. включающие частоты ассоциаций, подсчитанные отдельно для мужчин и для женщин.

Инструменты:
Flask (микрофреймворк для Python)
Bokeh (и другие библиотеки Python для визуализаций)
JavaScript

3) ЭТАПЫ

1. ноябрь, декабрь - создание опроса, чтение логов, подведение итогов опроса, прохождение tutorial, запуск у себя webvectors
2. январь - разбор кода, изучение библиотек, прохождение курсов по визуализации, имплементация
    a. фильтры по частотности
    b. фильтры по частям речи
3. февраль-март - создание динамического графа, начало написания статьи
4. апрель - завершение написания статьи и, если есть возможность, создание большой семантической карты и поиска по ней по семантическим классам НКРЯ
5. май - отладка фильтров и графов (и если есть возможность, создание большой семантической карты и поиска по ней по семантическим классам НКРЯ)
6. июнь - выкат всего нового функционала на сайт.

4) ТАБЛИЦА

	этап	Настя	Тома	Надя
январь	1	Ответственная по фильтрам ч, чр и (после прикрепления меток) НКРЯ. 1.Разбор кода webvectors 2.Фильтры по ч, чр: фронт 3.Изучение задачи по графам, подготовка плана декомпозиции	1.Разбор кода webvectors 2. Изучение и освещение основ UX и UI для оформления фронта 3.Прикрепление меток НКРЯ к словам	1.Разбор кода webvectors 2.Фильтры по ч, чр и (после прикрепления меток) НКРЯ: бэк 3.Изучение задачи по графам, подготовка плана декомпозиции

февраль	2	1.Имплементация части (1/3) плана по графам	Ответственный по графам. 1.Имплементация части (1/3) плана по графам	1.Имплементация части (1/3) плана по графам
март	3	1.Эксперименты с графами, отладка	1.Эксперименты с графами, отладка.	1.Эксперименты с графами, отладка.
апрель	4.	1.Написание статьи 2.Имплементация части (1/3) плана по сем. карте	1.Написание статьи 2.Имплементация части (1/3) плана по сем. карте	1.Написание статьи (Если успевается: Ответственный за создание семантической карты с фильтрами) 2.Имплементация части (1/3) плана по сем. карте
май	5.	1.Написание статьи 2. Доделывание функционала	1.Написание статьи 2. Доделывание функционала	1.Написание статьи 2. Доделывание функционала
июнь	6.	подготовка к выкату	подготовка к выкату	подготовка к выкату