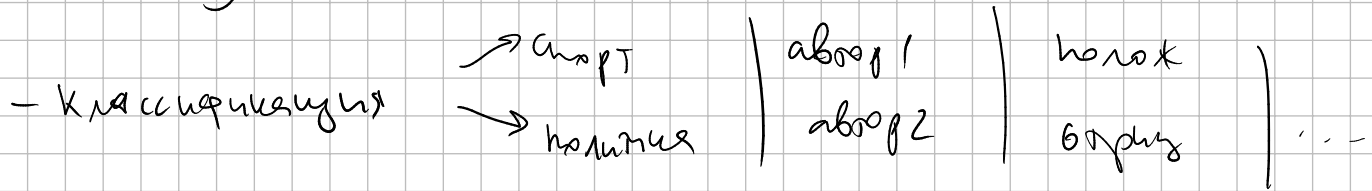


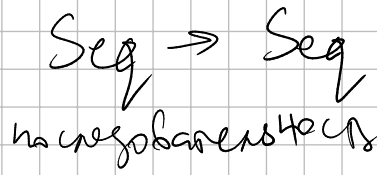
# Обработка ESI

- теги
  - рефы
- } иех. гоме



- классификация - см. классификация, но классификация аборт.

- абсолютн. перебор



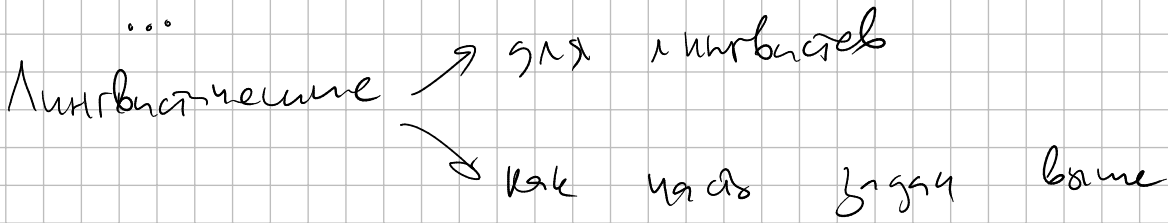
аборт переупорядочивание.

Длинный → короткий

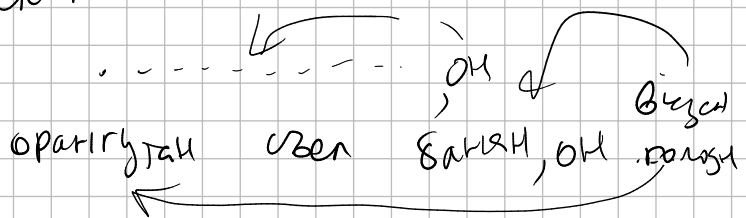
- итер. поиск

(поиск в базе рецептов)

...



- определение идент. сущностей
- размещение координат



- анализ значимости

- коллокации

- предикативность слов,

- ключевые слова.

Методы

I 1. Rule-based

методы, основанные на правилах

Перевод.

I have a dog

I

у меня есть

я генеро

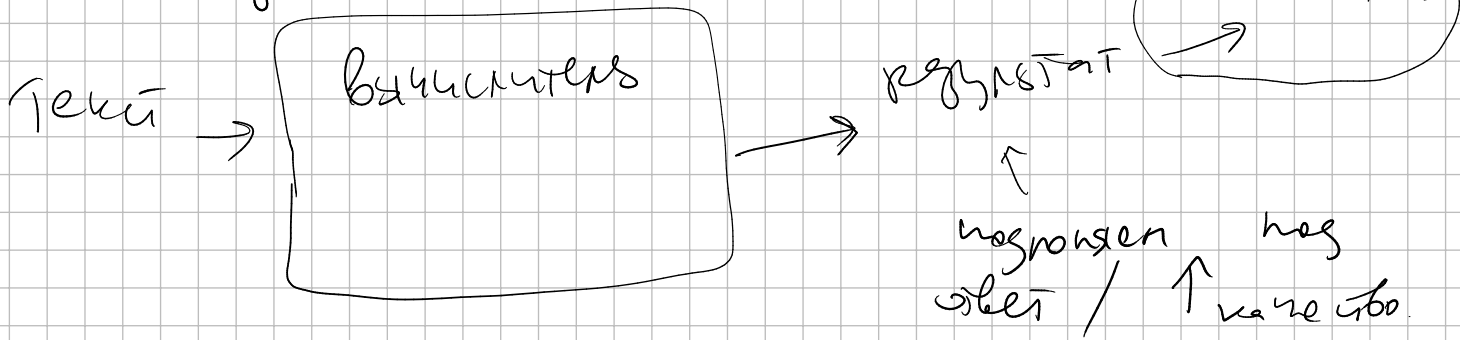
проблема: их много рассуждают, их много.

## II Статистические методы

Даны текст — считаем разное количество слов, вычисляем обрат (tf-idf)

Терминная оценка — термин бер-т, статистика, термин инф.

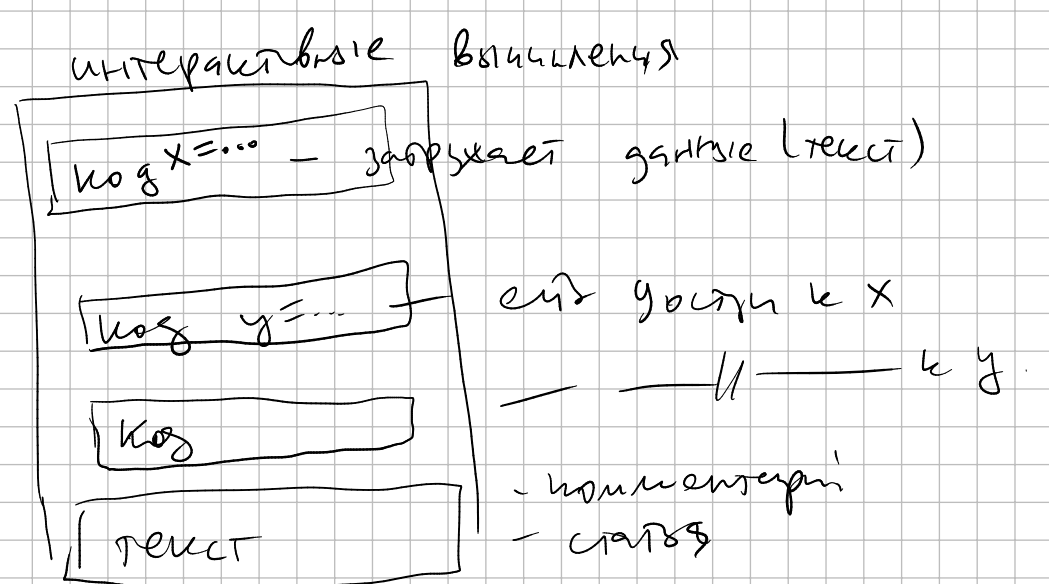
## III Маш. обучение.

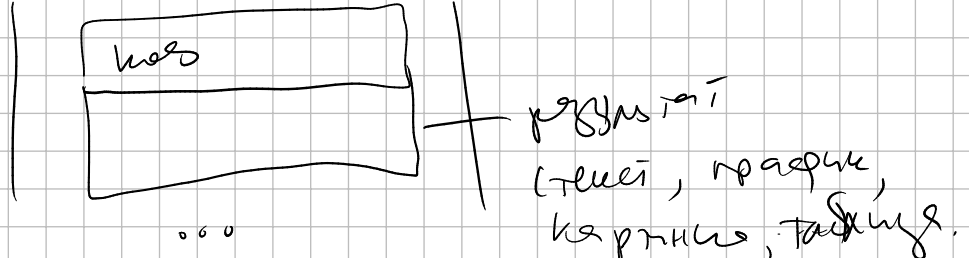


Python (nltk, numpy, pandas, scikit-learn) — язык — таблицы

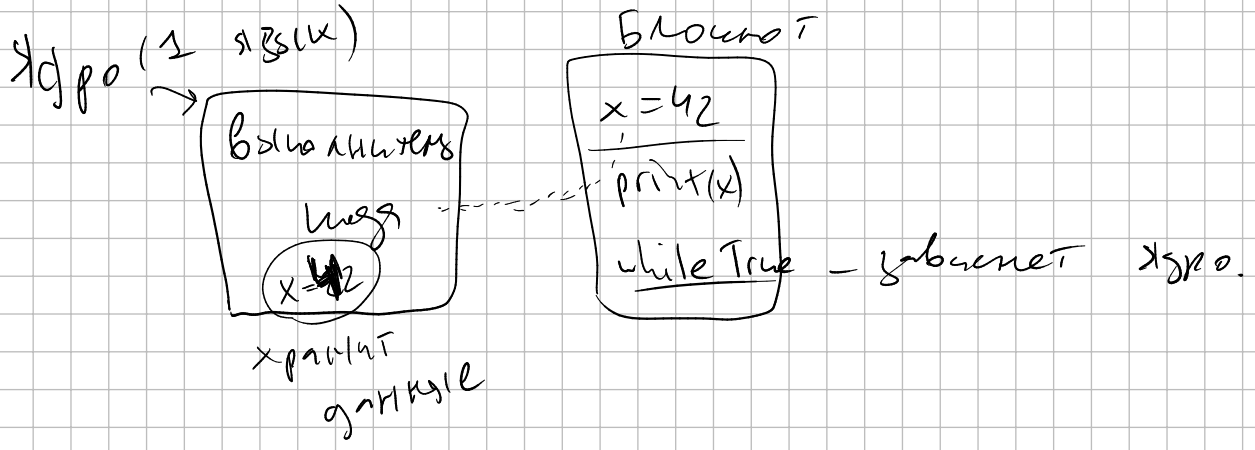
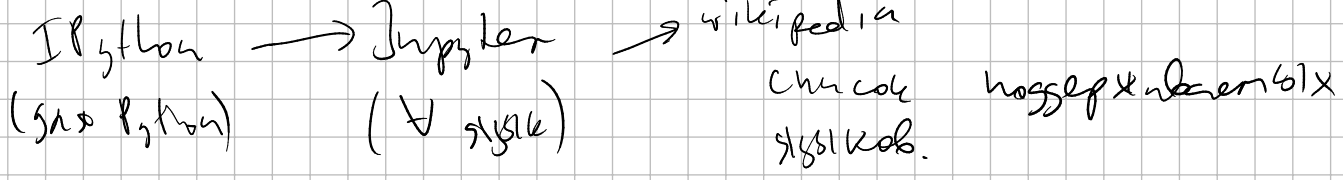
Jupyter  
Jupyter.

бандиты





Jupyter - та же технология блокнотов. еще другие.

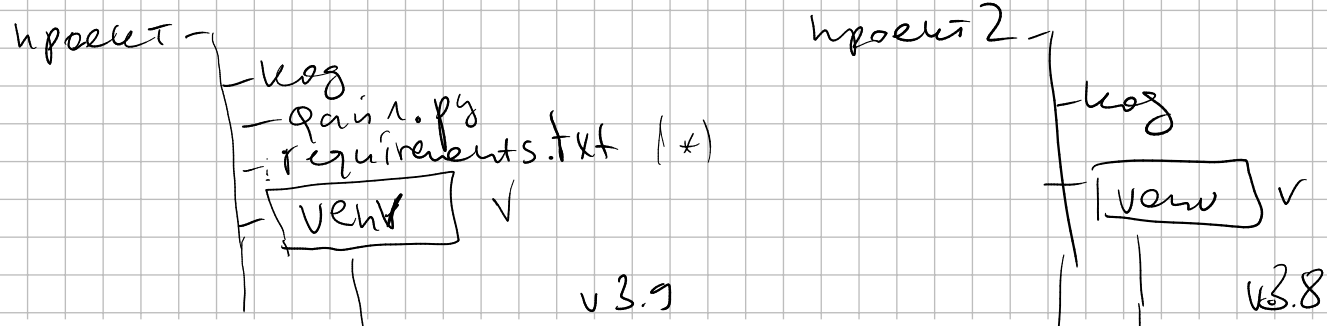


это может зависеть, зависит от того, что код записывается / через анкету / через статью, видео...

Использование.

- 1. У себя
  - установка в свой Python
  - Anaconda (устанавливать можно)
    - проще работать с библиотеками
    - граф. интерфейс

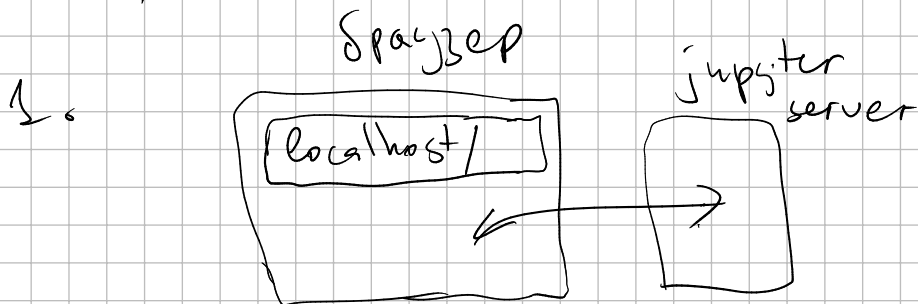
в проекте x к Python рекомендуете использовать вирт. окружения.



python, etc  
 системам.  
 http 3.5

python ce  
 http.  
 http 3.5

Установка.



1) в anaconda

- ein notebook

Jupyter Notebook

или

JupyterLab

2) из ком. строк

2. IntelliJ IDEA Ultimate  
 PyCharm

- работает лучше в IDEA.

Курсовая по 11:50

2. <sup>4</sup> гарантированные варианты. Изучить

- Google collaborative
- JetBrains datalore
- ...

Почин работа \* и проб.

- проект с темой + результативная

⇒ не нужно забывать про нас в процессе.

во всех интервалах их можно эксп/импортировать

- Просто: github - по запросу ipynb

статистика сервисов - ...

проект на github для связи загрузки

## Задача Закон Ципфа.

Английский текст, считаем (по убыванию)	частота rank	числ. frequency (частота)
и - 1000	1	1000
он - 500	2	500
кот - 330	3	333
:	4	250
:	5	200
:	:	:

Закон: frequency  $\sim \frac{1}{rank}$  ЭС: freq<sup>(1)</sup> =  $\frac{C}{rank}$



log (2)  $\log freq = \log \frac{C}{rank}$

