**Дорожкин Денис, 425 группа**

**Отчет**

3 исходных фразы

1) Нерелигиозный художник предпочёл написать солнце вместо Бога.

2) Два композитора, основоположники азербайджанской симфонической музыки, не только дружили, но и родились в один день.

3) Ерундопель хорош под водку.

Преобразованные морфологическим анализатором

1) нерелигиозный художник предпочитать написать солнце вместо бог

2) два композитор основоположник азербайджанский симфонический музыка не только дружить но и родиться в один день

3) ерундопель хороший под водка

Предобработка

Для каждой фразы были взяты все статьи на которую она ведет.

Статьи были скачаны, преобразованы анализатором и разбиты на фразы (документы) - каждая в отдельный файл.

Были сформированы списки файлов документов all1.txt all2.txt all3.txt

Программа search\_engine.py на вход принимает:

1) Имя файла с названиями всех файлов с фразами (документов)

2) Флаг - нужно ли использовать tf-idf или же только tf для вычисления весов

3) Имя файла с запросом

Все файлы уже обработаны морфологическим анализатором.

Далее программа вычисляет веса и выводит в порядке убывания пары (вес, имя документа).

Работа программы для каждой из фраз

1) нерелигиозный художник предпочитать написать солнце вместо бог

tf:

'morph2/1\_b\_50.txt', 0.30304576336566325

бог отец бог сын и бог дух святой

'morph2/1\_s\_100.txt', 0.26726124191242434

строение солнце

'morph2/1\_s\_156.txt', 0.24743582965269675

фотосфера образовывать видимый поверхность солнце по который определяться размер солнце расстояние от солнце и т д

'morph2/1\_s\_308.txt', 0.24174688920761409

благодаря спектроскопия быть обнаруживать новый элемент в состав солнце который называть гелий в честь древнегреческий бог солнце гелиос

'morph2/1\_p\_55.txt', 0.24174688920761409

данный работа вероятно быть написать под впечатление от наблюдение уоттса за звезда солнце через телескоп джеймс саута и отражать увлечение художник астрономия

tf-idf:

'morph2/1\_h\_19.txt', 0.19192927712151509

техника уоттса резко отличаться от техника большинство художник тот время предпочитать заглаживать живописный поверхность

'morph2/1\_p\_55.txt', 0.12957216356871795

данный работа вероятно быть написать под впечатление от наблюдение уоттса за звезда солнце через телескоп джеймс саута и отражать увлечение художник астрономия

'morph2/1\_p\_35.txt', 0.12158499634585644

картина солнце написать масло на холст а ее размер составлять см

'morph2/1\_n\_19.txt', 0.10931461924385792

в приводить ниже таблица по убывание показывать доля нерелигиозный часть население страна

'morph2/1\_p\_1.txt', 0.10192255802087748

картина написать в форма стилизованный морской пейзаж преимущественно яркий цветок

2) два композитор основоположник азербайджанский симфонический музыка не только дружить но и родиться в один день

tf:

'morph2/2\_d\_84.txt', 0.35424595421603805

не обходить композитор сторона и опыт европейский и в первый очередь русский композитор сплетать он с народный музыка и выводить такой образ последний на мировой уровень

'morph2/2\_k\_224.txt', 0.34426518632954811

в предисловие к основа азербайджанский народный музыка композитор писать

'morph2/2\_d\_1.txt', 0.33541019662496846

абдул муслим магомет оглы магомаев сентябрь грозный терский область российский империя июль нальчик ссср азербайджанский советский композитор и дирижер заслуживать деятель искусство азербайджанский сср один из основоположник азербайджанский классический музыка его имя называть азербайджанский государственный филармония

'morph2/2\_k\_61.txt', 0.32659863237109038

виноградов в свой книга узеир гаджибеков и азербайджанский музыка писать

'morph2/2\_k\_236.txt', 0.3227486121839514

но в тот же время он продолжать считать что азербайджанский музыка не должный оставаться всегда монодичной

tf-idf:

'morph2/2\_k\_218.txt', 0.20313935448329154

однако узеир гаджибеков известный на родина не только как композитор и меценат

'morph2/2\_s\_3.txt', 0.19323293095753777

с другой же сторона далеко не всякий музыка написать или перелагать для исполнение симфонический оркестр являться по свой внутренний свойство и качество музыка заведомо симфонический

'morph2/2\_s\_2.txt', 0.17530142129361342

хотя в большинство случай симфонический музыка предназначать для исполнение симфонический оркестр но то не менее симфонический по свой идейно содержательный значимость структура формообразование и специфика композиционный развитие мочь быть также и музыкальный произведение написать для какой-либо иной состав исполнитель как и для отдельный солировать инструмент или хор акапелла

'morph2/2\_d\_53.txt', 0.16725306254524719

за два день до представление быть убивать задействовать в постановка в качество режиссер гусейн араблинский но премьера состояться и быть удачный

'morph2/2\_a\_415.txt', 0.16222266472206179

при это рисовать не только декоративный мотив но и исторический портрет сцена охота и сражение

3) ерундопель хороший под водка

tf:

'morph2/3\_e\_7.txt', 0.44721359549995793

ерундопель подавать надо c водка

'morph2/3\_v\_77.txt', 0.40089186286863659

водка становиться подразделяться на водка и водка особый

'morph2/3\_v\_98.txt', 0.31980107453341566

предлагаться признавать исключительный право польша продавать и рекламировать на внешний рынок под имя водка свой товар а именно водка выборову

'morph2/3\_v\_113.txt', 0.31622776601683794

только водка из россия быть настоящий русский водка

'morph2/3\_v\_37.txt', 0.29999999999999999

основываться на этот данные менделеев разрабатывать рецептура водка под название московский особенный который в год быть запатентовывать российский правительство как русский национальный водка

tf-idf:

'morph2/3\_e\_7.txt', 0.19717345336205047

ерундопель подавать надо c водка

'morph2/3\_e\_8.txt', 0.17393751475970001

ерундопель спрашивать удивленно это пирожок

'morph2/3\_e\_1.txt', 0.14724827445102825

ерундопель закусочный салат из икра рыба и овощ

'morph2/3\_e\_24.txt', 0.13379868945617449

сборник рассказ татьяна михайловская называться ерундопель и другой

'morph2/3\_v\_37.txt', 0.11278761332812513

основываться на этот данные менделеев разрабатывать рецептура водка под название московский особенный который в год быть запатентовывать российский правительство как русский национальный водка