**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 lang1.txt lang2.txt lang3.txt

Программа language\_model.py на вход принимает:

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

2) Имя файла с запросом

Все файлы уже обработаны морфологическим анализатором.

Мера nDCG для каждого запроса считалась для 2х значений лямбда (0.5 и 0.9)

Программа вычисляет веса и выводит в порядке убывания пары (вес, имя документа) по известной формуле.

также программа вычисляет DCG, iDCG, nDCG метрику и выводит ее.

Работа программы для каждой из фраз (выписаны первые 5 результатов)

1. нерелигиозный художник предпочитать написать солнце вместо бог

**Проставление оценок экспертом**

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

**оценка 2**

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

**оценка 1**

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

**оценка 1**

**Далее оценки 0**

**l = 0.5:**

morph3/1\_h\_19.txt 2.63126477224e-19

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

morph3/1\_p\_55.txt 7.21128043348e-20

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

morph3/1\_p\_35.txt 1.06339060603e-20

картина солнце написать масло на холст а ее размер составлять сто см

morph3/1\_p\_45.txt 4.40242988349e-21

я не пытаться написать портрет солнце такой не нарисовать но я хотеть производить впечатление передавать его колоссальный сила

morph3/1\_p\_49.txt 2.88669939997e-21

например на картина жертвоприношение ной написать Уоттсом в год видно как ной приносить жертва солнце в благодарность за спасение свой семья от потоп

DCG: 2.4229042789

iDCG: 3.13092975357

nDCG: 0.77386095173

**l = 0.9:**

morph3/1\_p\_55.txt 4.80078985718e-22

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

morph3/1\_h\_19.txt 2.65556853387e-22

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

morph3/1\_p\_35.txt 9.37450262353e-24

картина солнце написать масло на холст а ее размер составлять сто см

morph3/1\_p\_45.txt 3.54967285509e-24

я не пытаться написать портрет солнце такой не нарисовать но я хотеть производить впечатление передавать его колоссальный сила

morph3/1\_p\_49.txt 2.19541852233e-24

например на картина жертвоприношение ной написать Уоттсом в год видно как ной приносить жертва солнце в благодарность за спасение свой семья от потоп

DCG: 2.79197452533

iDCG: 3.13092975357

nDCG: 0.89173975307

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

**Проставление оценок экспертом**

1) 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   
**оценка 1**

2) 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   
**оценка 1**

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

**оценка 1**   
4) 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

**оценка 1**

5) 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

**оценка 1   
далее оценки 0**

**l=0.5:**

morph3/2\_d\_53.txt 9.46996269992e-41

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

morph3/2\_d\_1.txt 6.89774157284e-41

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

morph3/2\_a\_415.txt 6.61912463038e-41

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

morph3/2\_k\_218.txt 4.21872230166e-41

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

morph3/2\_k\_236.txt 2.27880022059e-41

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

DCG: 2.05217682775

iDCG: 2.94845911888

nDCG: 0.696016714158

**l=0.9:**

morph3/2\_d\_1.txt 3.93296662901e-46

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

morph3/2\_d\_53.txt 6.95617751675e-47

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

morph3/2\_s\_2.txt 6.0090620942e-47

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 а капелла

morph3/2\_k\_236.txt 1.86828775279e-47

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

morph3/2\_a\_415.txt 1.07536625989e-47

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

DCG: 2.49057051611

iDCG: 2.94845911888

nDCG: 0.844702407491

3) ерундопель хороший под водка

**Проставление оценок экспертом**

1) ерундопель подавать надо c водка

**оценка 2**

2) ерундопель закусочный салат из икра рыба и овощ

**оценка 1**

**далее оценки 0**

**l=0.5:**

morph3/3\_e\_7.txt 1.63815542233e-09

Ерундопель подавать надо c водка

morph3/3\_e\_8.txt 2.45152542842e-10

Ерундопель спрашивать удивленно пирожок

morph3/3\_v\_98.txt 2.07304908435e-10

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

morph3/3\_v\_37.txt 1.62783756294e-10

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

morph3/3\_v\_114.txt 1.49279173087e-10

страна водочный пояс входить в ес энергично противодействовать попытка продавать под название водка изделие на основа виноградный сырье

DCG: 2.33333333333

iDCG: 2.63092975357

nDCG: 0.886885455671

**l=0.9:**

morph3/3\_e\_7.txt 1.88107426342e-10

Ерундопель подавать надо c водка

morph3/3\_v\_98.txt 2.09585889653e-11

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

morph3/3\_v\_37.txt 1.59288741284e-11

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

morph3/3\_v\_114.txt 1.32353909072e-11

страна водочный пояс входить в ес энергично противодействовать попытка продавать под название водка изделие на основа виноградный сырье

morph3/3\_v\_228.txt 1.16374591968e-11

в москва быть музей история водка который до октябрь год располагаться в санкт-петербург в год после реконструкция в санкт-петербург открываться обновлять экспозиция под название музей русский водка

DCG: 2.30102999566

iDCG: 2.63092975357

nDCG: 0.874607158378

**Среднее nDCG для языковой модели с l = 0.5:**

0.7855877071863334

**Среднее nDCG для языковой модели с l = 0.9:**

0.8703497729796666

**Среднее nDCG для векторной модели с tf:**

0.801282784613

**Среднее nDCG для векторной модели с tf-idf:**

0.936651038881

**Оценка nDCG**

nDCG(l = 0.5) < nDCG(tf) < nDCG(l = 0.9) < nDCG(tf-idf)

**Оценка моделей (от лучшей к худшей по мере nDCG)**

Векторная модель (tf-idf)

Языковая модель (l = 0.9)

Векторная модель (tf)

Языковая модель (l = 0.5)