

IR ASSIGNMENT 2

Question 1)

```
In [1]: runfile('/home/kaushal/Desktop/IR_A2_Q1.py', wdir='/home/kaushal/Desktop')

Enter query
"G00d DaY"
19
mazarin.txt
enchdup.hum
sick-kid.txt
brain.damage
aesop11.txt
hound-b.txt
startrek.txt
srex.txt
fantasy.txt
bruce-p.txt
fantasy.hum
superg1
aesopa10.txt
melissa.txt
horsewolf.txt
outcast.dos
forgotte
fic5
13chil.txt
```

Question 2)

Jaccard Coefficient

```
In [1]: runfile('/home/kaushal/Desktop')
Jaccard Coefficient

Enter query
"The crab and the heron"
crabhern.txt
index.html
missing.txt
aesopa10.txt
aesop11.txt
```

TF-IDF Schemes :

```
Please enter query
"The crab and the heron"
Using Scheme 1
crabhern.txt
aesop11.txt
index.html
fgoose.txt
aesopa10.txt

Using Scheme 2
crabhern.txt
aesop11.txt
timem.hac
aesopa10.txt
long1-3.txt

Using Scheme 3
crabhern.txt
aesopa10.txt
long1-3.txt
aesop11.txt
missing.txt

: ready  Kite: ready  cu:
```

```
Using Scheme 4
crabhern.txt
aesop11.txt
timem.hac
index.html
aesopa10.txt
```

```
Using Scheme 5
crabhern.txt
aesop11.txt
aesopa10.txt
long1-3.txt
missing.txt
```

Cosine Similarity:

```
Please enter query
The crab and the heron
Using Scheme 1
crabhern.txt
index.html
aesop11.txt
missing.txt
aesopa10.txt

Using Scheme 2
crabhern.txt
aesopa10.txt
aesop11.txt
timem.hac
long1-3.txt

Using Scheme 3
crabhern.txt
aesopa10.txt
aesop11.txt
timem.hac
long1-3.txt
```

```
Using Scheme 4
crabhern.txt
aesop11.txt
aesopa10.txt
long1-3.txt
index.html
```

```
Using Scheme 5
crabhern.txt
aesop11.txt
aesopa10.txt
long1-3.txt
missing.txt
```

Question 3)

```
In [1]: runfile('/home/kaushal/Desktop/IR_A2_Q
Max DCG is 20.989750804831445
```

Possible order of query-url pair with MAX DCG

8

19

20

22

23

26

35

37

38

41

53

59

62

63

69

77

91

101

5

7

11

47

Total no of files possible with max DCG are

1989349737593837059982604761490532989693684017056657058820518031270485799269519348241268656543

10502400000000000000000000000000

```
In [1]: runfile('/home/kaushal/Desktop/IR_A2_Q3_c.py', wdir='/home/kaushal/Desktop')
```

Normalized DCG for first 50 docs is 0.3521042740324887

Normalized DCG for whole dataset is 0.5979226516897831

```
In [2]:
```

```
In [1]: runfile('/home/kaushal/Desktop/IR_A2_Q3_c.py', wdir='/home/kaushal/Desktop')
```

