**To evaluate retrieval effectiveness**

**Setup, Compilation, Execution and Output Guidelines**

This is an implementation code designed in python 2.7 for evaluating small search engine (Search Engine Evaluated is Lucene Implementation- HW4)

For this code to be executed on any terminal, we need to have below pre-requisites met:

1. Python 2.7 installed

Go to https://www.python.org/downloads/ -> install as per operating system

Required Python Packages-

from string import lstrip  
from collections import OrderedDict  
import sys  
import re  
import math

Requirement:

File name: evaluation.py

No. of Input Arguments: Two

1. lucene\_output.txt

This file in the input argument containing already stemmed document collection

1. cacm.rel

This file contains the relevance judgement for CACM test collection

Steps to be done:

Place the given python file i. e. “evaluation.py” along with lucene search engine output file “lucene\_output.txt” and relevance judgement file “cacm.rel” to any particular location (File path)

Compilation and Run Python code:

Go to terminal

Type python. This will start the python shell terminal

Navigate to the required path and run .py file as below->

->FilePath/ evaluation.py lucene\_output.txt cacm.rel

Expected Output:

Console Output:

Prints precision at rank 20 i.e. P@20 for given 3 queries

1 -> portable operating systems = 0.2

2 -> code optimization for space efficiency = 0.2

3 -> parallel algorithms = 0.35

Prints the count of relevant documents for each of the queries as per the relevance judgement.

Displays the average precision value calculated for each query and then the Mean Average Precision value i.e. MAP = 0.43326171222

NOTE: Above statistical console output is present in file -> Effectiveness\_op\_3\_queries.txt

Also, the query ids for lucene\_output file and the given CACM relevance judgement is mapped as below:

Query 1 <-> lucene\_output -> Query 1 <-> CACM Rel file -> Query 12

Query 2 <-> lucene\_output -> Query 2 <-> CACM Rel file -> Query 13

Query 3 <-> lucene\_output -> Query 3 <-> CACM Rel file -> Query 19

File Output:

As required, creates 3 output files corresponding to table for each of the given three queries ->

Query 1 -> table\_1.txt

Query 2 -> table\_2.txt

Query 3 -> table\_3.txt

Containing below details for top 100 documents for each query

Rank

Document ID

Document score

Relevance level

Precision

Recall

NDCG

REFERENCES:

1) http://www.tutorialspoint.com/

2) <https://www.youtube.com/>

3) www.stackoverflow.com/

----------------------------------------------END---------------------------------------------------------