**SECTION 1:**

----------------------------------------------------------------

Starting process for Step1

----------------------------------------------------------------

Building index for step: Step1

Writing Inverted index to file for Step1

**Get Index statistics for Step1**

-----------------------------------------------

Number of Terms: 15991

Maximum Length of Postings List:501

Minimum Length of Postings List:1

Average Length of Postings List:10.454630729785505

Size of the file that stores the inverted index: 993.685546875 KBs

----------------------------------------------------------------

Starting process for Step2

----------------------------------------------------------------

Remove stop words and small words for Step2

Building index for step: Step2

Writing Inverted index to file for Step2

**Get Index statistics for Step2**

-----------------------------------------------

Number of Terms: 15575

Maximum Length of Postings List:501

Minimum Length of Postings List:1

Average Length of Postings List:7.309020866773675

Size of the file that stores the inverted index: 739.1748046875 KBs

----------------------------------------------------------------

Starting process for Step3

----------------------------------------------------------------

Remove stop words and small words for Step3

Applying Porter Stemmer on tokens for Step3

Building index for step: Step3

Getting the plot of Terms v/s Tokens for each document for Step3

This long task might take few minutes

Writing Inverted index to file for Step3

**Get Index statistics for Step3**

-----------------------------------------------

Number of Terms: 10932

Maximum Length of Postings List:501

Minimum Length of Postings List:1

Average Length of Postings List:9.65642151481888

Size of the file that stores the inverted index: 632.283203125 KBs

----------------------------------------------------------------

Starting process for Step4

----------------------------------------------------------------

Remove stop words and small words for Step4

Applying Porter Stemmer on tokens for Step4

Building index for step: Step4

Removing least frequent terms for Step4

Writing Inverted index to file for Step4

**Get Index statistics for Step4**

-----------------------------------------------

Number of Terms: 2892

Maximum Length of Postings List:501

Minimum Length of Postings List:6

Average Length of Postings List:31.467842323651453

Size of the file that stores the inverted index: 464.181640625 KBs

**SECTION 2:**

**Printing stats for most frequent 20 words**

----------------------------------------------------------------

For K = 1 and token = UK

Postings list size:501

Average Gap Size in the Postings List: 1.0

-----------------------------

For K = 2 and token = ski

Postings list size:501

Average Gap Size in the Postings List: 1.0

-----------------------------

For K = 3 and token = updat

Postings list size:496

Average Gap Size in the Postings List: 1.0101010101010102

-----------------------------

For K = 4 and token = nnew

Postings list size:490

Average Gap Size in the Postings List: 1.0224948875255624

-----------------------------

For K = 5 and token = the

Postings list size:456

Average Gap Size in the Postings List: 1.0967032967032968

-----------------------------

For K = 6 and token = nlast

Postings list size:454

Average Gap Size in the Postings List: 1.1015452538631347

-----------------------------

For K = 7 and token = time

Postings list size:394

Average Gap Size in the Postings List: 1.2697201017811706

-----------------------------

For K = 8 and token = win

Postings list size:385

Average Gap Size in the Postings List: 1.3020833333333333

-----------------------------

For K = 9 and token = It

Postings list size:348

Average Gap Size in the Postings List: 1.440922190201729

-----------------------------

For K = 10 and token = final

Postings list size:344

Average Gap Size in the Postings List: 1.4285714285714286

-----------------------------

For K = 11 and token = play

Postings list size:344

Average Gap Size in the Postings List: 1.4373177842565599

-----------------------------

For K = 12 and token = game

Postings list size:332

Average Gap Size in the Postings List: 1.498489425981873

-----------------------------

For K = 13 and token = sport

Postings list size:326

Average Gap Size in the Postings List: 1.5353846153846153

-----------------------------

For K = 14 and token = look

Postings list size:312

Average Gap Size in the Postings List: 1.607717041800643

-----------------------------

For K = 15 and token = player

Postings list size:299

Average Gap Size in the Postings List: 1.6543624161073827

-----------------------------

For K = 16 and token = top

Postings list size:289

Average Gap Size in the Postings List: 1.7222222222222223

-----------------------------

For K = 17 and token = won

Postings list size:288

Average Gap Size in the Postings List: 1.7386759581881532

-----------------------------

For K = 18 and token = follow

Postings list size:288

Average Gap Size in the Postings List: 1.735191637630662

-----------------------------

For K = 19 and token = He

Postings list size:284

Average Gap Size in the Postings List: 1.7667844522968197

-----------------------------

For K = 20 and token = match

Postings list size:282

Average Gap Size in the Postings List: 1.7615658362989324

-----------------------------

----------------------------------------------------------------

**Printing stats for least frequent 20 words**

----------------------------------------------------------------

For K = 1 and token = republ

Postings list size:6

Average Gap Size in the Postings List: 18.6

-----------------------------

For K = 2 and token = outfield

Postings list size:6

Average Gap Size in the Postings List: 45.4

-----------------------------

For K = 3 and token = struff

Postings list size:6

Average Gap Size in the Postings List: 83.0

-----------------------------

For K = 4 and token = investig

Postings list size:6

Average Gap Size in the Postings List: 50.6

-----------------------------

For K = 5 and token = hunt

Postings list size:6

Average Gap Size in the Postings List: 77.6

-----------------------------

For K = 6 and token = son

Postings list size:6

Average Gap Size in the Postings List: 44.4

-----------------------------

For K = 7 and token = jon

Postings list size:6

Average Gap Size in the Postings List: 41.4

-----------------------------

For K = 8 and token = rafal

Postings list size:6

Average Gap Size in the Postings List: 13.4

-----------------------------

For K = 9 and token = truth

Postings list size:6

Average Gap Size in the Postings List: 44.4

-----------------------------

For K = 10 and token = brink

Postings list size:6

Average Gap Size in the Postings List: 81.8

-----------------------------

For K = 11 and token = assert

Postings list size:6

Average Gap Size in the Postings List: 87.4

-----------------------------

For K = 12 and token = diminish

Postings list size:6

Average Gap Size in the Postings List: 57.2

-----------------------------

For K = 13 and token = yorkshir

Postings list size:6

Average Gap Size in the Postings List: 96.2

-----------------------------

For K = 14 and token = gaal

Postings list size:6

Average Gap Size in the Postings List: 46.8

-----------------------------

For K = 15 and token = restor

Postings list size:6

Average Gap Size in the Postings List: 75.4

-----------------------------

For K = 16 and token = pin

Postings list size:6

Average Gap Size in the Postings List: 63.8

-----------------------------

For K = 17 and token = partli

Postings list size:6

Average Gap Size in the Postings List: 61.8

-----------------------------

For K = 18 and token = chatrier

Postings list size:6

Average Gap Size in the Postings List: 73.2

-----------------------------

For K = 19 and token = thomaz

Postings list size:6

Average Gap Size in the Postings List: 43.0

-----------------------------

For K = 20 and token = instinct

Postings list size:6

Average Gap Size in the Postings List: 52.6

-----------------------------

----------------------------------------------------------------

**Printing stats for Median frequent 20 words**

----------------------------------------------------------------

For K = 1436 and token = hero

Postings list size:15

Average Gap Size in the Postings List: 17.642857142857142

-----------------------------

For K = 1437 and token = haul

Postings list size:15

Average Gap Size in the Postings List: 27.285714285714285

-----------------------------

For K = 1438 and token = spoke

Postings list size:15

Average Gap Size in the Postings List: 27.571428571428573

-----------------------------

For K = 1439 and token = assur

Postings list size:15

Average Gap Size in the Postings List: 30.357142857142858

-----------------------------

For K = 1440 and token = smith

Postings list size:15

Average Gap Size in the Postings List: 30.428571428571427

-----------------------------

For K = 1441 and token = know

Postings list size:15

Average Gap Size in the Postings List: 27.928571428571427

-----------------------------

For K = 1442 and token = easili

Postings list size:15

Average Gap Size in the Postings List: 33.57142857142857

-----------------------------

For K = 1443 and token = blister

Postings list size:15

Average Gap Size in the Postings List: 32.07142857142857

-----------------------------

For K = 1444 and token = bristol

Postings list size:15

Average Gap Size in the Postings List: 29.571428571428573

-----------------------------

For K = 1445 and token = refere

Postings list size:15

Average Gap Size in the Postings List: 30.928571428571427

-----------------------------

For K = 1446 and token = four

Postings list size:15

Average Gap Size in the Postings List: 28.0

-----------------------------

For K = 1447 and token = alan

Postings list size:15

Average Gap Size in the Postings List: 31.357142857142858

-----------------------------

For K = 1448 and token = gareth

Postings list size:15

Average Gap Size in the Postings List: 31.928571428571427

-----------------------------

For K = 1449 and token = exchang

Postings list size:15

Average Gap Size in the Postings List: 34.642857142857146

-----------------------------

For K = 1450 and token = nearli

Postings list size:15

Average Gap Size in the Postings List: 24.142857142857142

-----------------------------

For K = 1451 and token = reject

Postings list size:15

Average Gap Size in the Postings List: 31.0

-----------------------------

For K = 1452 and token = 09/08/17

Postings list size:15

Average Gap Size in the Postings List: 14.5

-----------------------------

For K = 1453 and token = composur

Postings list size:15

Average Gap Size in the Postings List: 27.428571428571427

-----------------------------

For K = 1454 and token = lower

Postings list size:15

Average Gap Size in the Postings List: 32.42857142857143

-----------------------------

For K = 1455 and token = benefit

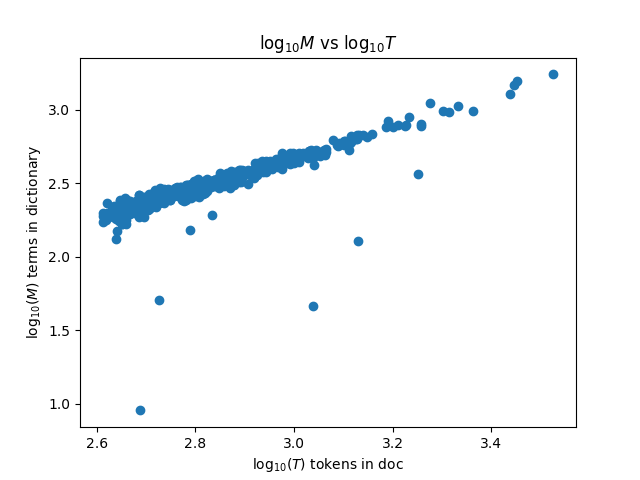
Postings list size:15

Average Gap Size in the Postings List: 28.071428571428573

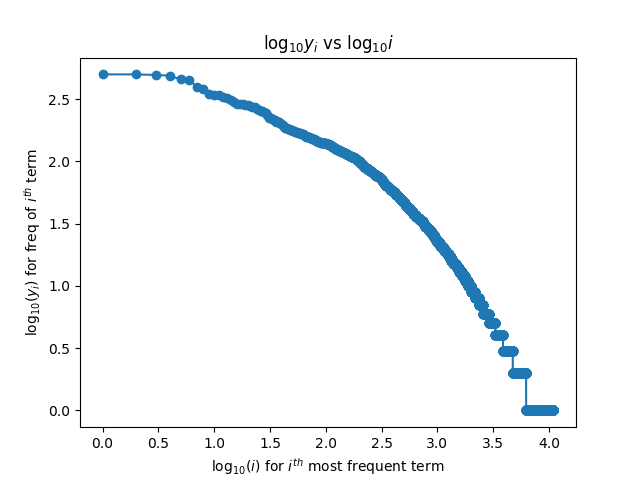
-----------------------------

----------------------------------------------------------------

**SECTION 3:**

1.

2.



3.

