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

Postings list size:501

Average Gap Size in the Postings List: 1.0

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

For K = 2

Postings list size:501

Average Gap Size in the Postings List: 1.0

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

For K = 3

Postings list size:496

Average Gap Size in the Postings List: 1.0101010101010102

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

For K = 4

Postings list size:490

Average Gap Size in the Postings List: 1.0224948875255624

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

For K = 5

Postings list size:456

Average Gap Size in the Postings List: 1.0967032967032968

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

For K = 6

Postings list size:454

Average Gap Size in the Postings List: 1.1015452538631347

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

For K = 7

Postings list size:394

Average Gap Size in the Postings List: 1.2697201017811706

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

For K = 8

Postings list size:385

Average Gap Size in the Postings List: 1.3020833333333333

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

For K = 9

Postings list size:348

Average Gap Size in the Postings List: 1.440922190201729

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

For K = 10

Postings list size:344

Average Gap Size in the Postings List: 1.4285714285714286

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

For K = 11

Postings list size:344

Average Gap Size in the Postings List: 1.4373177842565599

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

For K = 12

Postings list size:332

Average Gap Size in the Postings List: 1.498489425981873

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

For K = 13

Postings list size:326

Average Gap Size in the Postings List: 1.5353846153846153

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

For K = 14

Postings list size:312

Average Gap Size in the Postings List: 1.607717041800643

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

For K = 15

Postings list size:299

Average Gap Size in the Postings List: 1.6543624161073827

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

For K = 16

Postings list size:289

Average Gap Size in the Postings List: 1.7222222222222223

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

For K = 17

Postings list size:288

Average Gap Size in the Postings List: 1.735191637630662

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

For K = 18

Postings list size:288

Average Gap Size in the Postings List: 1.7386759581881532

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

For K = 19

Postings list size:284

Average Gap Size in the Postings List: 1.7667844522968197

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

For K = 20

Postings list size:282

Average Gap Size in the Postings List: 1.7615658362989324

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

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

**Printing stats for least frequent 20 words**

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

For K = 1

Postings list size:6

Average Gap Size in the Postings List: 68.6

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

For K = 2

Postings list size:6

Average Gap Size in the Postings List: 72.6

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

For K = 3

Postings list size:6

Average Gap Size in the Postings List: 73.2

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

For K = 4

Postings list size:6

Average Gap Size in the Postings List: 87.4

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

For K = 5

Postings list size:6

Average Gap Size in the Postings List: 57.8

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

For K = 6

Postings list size:6

Average Gap Size in the Postings List: 89.2

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

For K = 7

Postings list size:6

Average Gap Size in the Postings List: 87.2

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

For K = 8

Postings list size:6

Average Gap Size in the Postings List: 68.8

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

For K = 9

Postings list size:6

Average Gap Size in the Postings List: 44.2

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

For K = 10

Postings list size:6

Average Gap Size in the Postings List: 77.4

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

For K = 11

Postings list size:6

Average Gap Size in the Postings List: 50.6

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

For K = 12

Postings list size:6

Average Gap Size in the Postings List: 90.4

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

For K = 13

Postings list size:6

Average Gap Size in the Postings List: 67.2

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

For K = 14

Postings list size:6

Average Gap Size in the Postings List: 36.4

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

For K = 15

Postings list size:6

Average Gap Size in the Postings List: 68.0

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

For K = 16

Postings list size:6

Average Gap Size in the Postings List: 73.2

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

For K = 17

Postings list size:6

Average Gap Size in the Postings List: 4.6

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

For K = 18

Postings list size:6

Average Gap Size in the Postings List: 59.2

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

For K = 19

Postings list size:6

Average Gap Size in the Postings List: 39.6

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

For K = 20

Postings list size:6

Average Gap Size in the Postings List: 80.6

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

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

**Printing stats for Median frequent 20 words**

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

For K = 1436

Postings list size:15

Average Gap Size in the Postings List: 27.285714285714285

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

For K = 1437

Postings list size:15

Average Gap Size in the Postings List: 30.428571428571427

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

For K = 1438

Postings list size:15

Average Gap Size in the Postings List: 27.928571428571427

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

For K = 1439

Postings list size:15

Average Gap Size in the Postings List: 33.92857142857143

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

For K = 1440

Postings list size:15

Average Gap Size in the Postings List: 33.42857142857143

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

For K = 1441

Postings list size:15

Average Gap Size in the Postings List: 22.285714285714285

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

For K = 1442

Postings list size:15

Average Gap Size in the Postings List: 24.714285714285715

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

For K = 1443

Postings list size:15

Average Gap Size in the Postings List: 31.357142857142858

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

For K = 1444

Postings list size:15

Average Gap Size in the Postings List: 27.428571428571427

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

For K = 1445

Postings list size:15

Average Gap Size in the Postings List: 20.928571428571427

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

For K = 1446

Postings list size:15

Average Gap Size in the Postings List: 28.928571428571427

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

For K = 1447

Postings list size:15

Average Gap Size in the Postings List: 31.285714285714285

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

For K = 1448

Postings list size:15

Average Gap Size in the Postings List: 8.714285714285714

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

For K = 1449

Postings list size:15

Average Gap Size in the Postings List: 33.785714285714285

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

For K = 1450

Postings list size:15

Average Gap Size in the Postings List: 29.571428571428573

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

For K = 1451

Postings list size:15

Average Gap Size in the Postings List: 31.857142857142858

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

For K = 1452

Postings list size:15

Average Gap Size in the Postings List: 30.5

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

For K = 1453

Postings list size:15

Average Gap Size in the Postings List: 31.0

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

For K = 1454

Postings list size:15

Average Gap Size in the Postings List: 27.357142857142858

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

For K = 1455

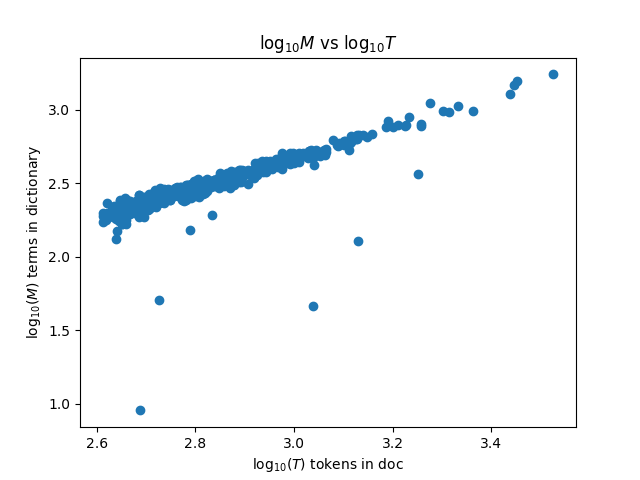
Postings list size:15

Average Gap Size in the Postings List: 32.357142857142854

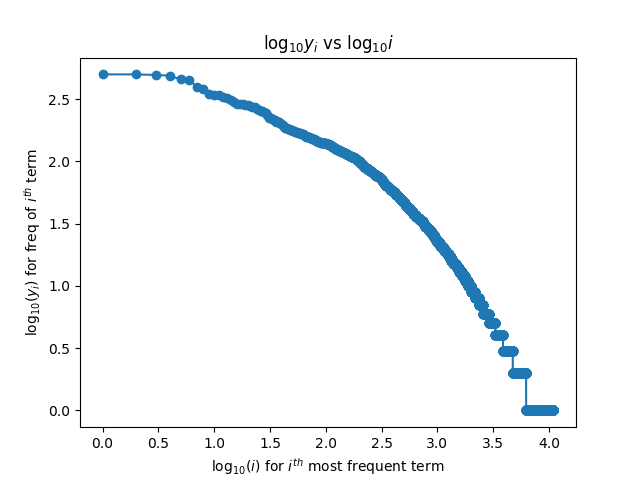
-----------------------------

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

SECTION 3:

1.

2.



3.