SUMMARIZER MANUAL

-by Ayush Pareek

INSTRUCTION MANUAL AND FAQs / DOCUMENTATION

Q1. How to run this program.

Ans. (Please use GCC C++ compiler 4.9.3 on Linux or Eclipse/Codeblocks on windows)

STEP 1::Put the text you want to summarize in "raw.txt"(sample text is already provided). Let the program run completely.

STEP 2::You will be asked the number of lines in which you want the summary. Give as any as you want (PCC>0.7 is not used fearing that the summary would become blank if none of the sentences qualify that criterion)

STEP 3::O/P will be displayed on the screen as well as on "summary.txt"

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

Q2. How to track the progress of the program to check its correctness?

Ans.

CHECK 1: You will see each sentence being separated into its own text file numbered as "<position of sentence >.txt".

Eg; 5.txt refers to the text file containing only the 5th sentence of the original text

CHECK 2: It has been made sure that pseudo full-stops (Eg; between acronyms like M.B.B.S. and after salutations like Mr. Mrs. Dr. Ms.) don't break the sentences. Put contents of 'raw2.txt' in 'raw.txt' to check the same

CHECK 3: Various text files will be formed. Their description is provided::

1)post\_formatting.txt:: GIVES each document after removing symbols and numbers and converting ' '(space) to '\n'(next line) character. This is done by the Format() function in the code.

2)post\_stemming.txt:: Gives each document after passing it through the Portal stemmer.(SEE:: PreStem() and stemfile() function)

3)post\_sorting.txt:: After lexicographical Sorting , only unique words are printed in the documents.

4)post\_stopword\_filtering.txt:: Sorted words without the stop words

CHECK 4: post\_stemming & post\_formatting are two folders which are not supposed to be deleted. They display each sentence file after doing the operation suggested by their names

CHECK 5: "matrix.txt" shows a WORD v/s SENTENCE MATRIX

CHECK 6: "coefficient\_matrix.txt" displays an N\*N matrix where each element i.e. c(i,j) is the Pearson Correlation Coefficient of Sentence i and Sentence j. You can see that whenever (i == j) i.e. for diagonal elements ==> c(i,j) == 1

CHECK 7:Finally the summary is displayed in summary.txt

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