Assignment 1 output.pdf

File used is: data/20 newsgroups/rec.motorcycles/103117

Task 1:

Number of sentences present are: 36 Number of tokens present are: 680 Number of words present are: 558

Number of unique words/tokens are: 327/343.

Task 2.

Number of words starting with consonants are: 321 Number of words starting with vowels are: 112

Number of unique words starting with consonants/vowels are: 200/46.

Task 3:

Number of email IDs present in the given text file are: 5 Number of UNIQUE email IDs present in the given text file are: 4 And the email ID are as follows:

manes@magpie.linknet.com C4tFwI.DLp@magpie.linknet.com 1pd49g\$a8m@vtserf.cc.vt.edu ranck@joesbar.cc.vt.edu manes@magpie.linknet.com

Task 4:

Number of sentences starting with --"you"-- are: 2 And the sentences are as follows:

You guessed it. You got off cheap.

Task 5:

Number of sentences ending with --"cheap."-- are: 1 And the sentences are as follows:

You got off cheap.

Task 5:

Number of sentences containing the word--"100"-- are: 1 Number of words containing the word--"100"-- are: 1 And the sentences are as follows:

100 miles in rain: and darkness.

Task 7:

Number of questions present are: 0 And the questions are as follows:

Task 8:

Mins and Secs present in the file are as follows: --43 mins, 29 sec

Task 9:

Number of abbreviations present in the file are: 15 Number of UNIQUE abbreviations present in the file are: 13 Abbreviations present in the file are as follows:

- --NYC
- --GMT
- --Message-ID
- --TIN
- --PL9
- --K75RT
- --DAMN
- --BMW
- --BMW
- **--**CT
- --K100RS
- --BMW
- --FLHS
- --NY
- --USA