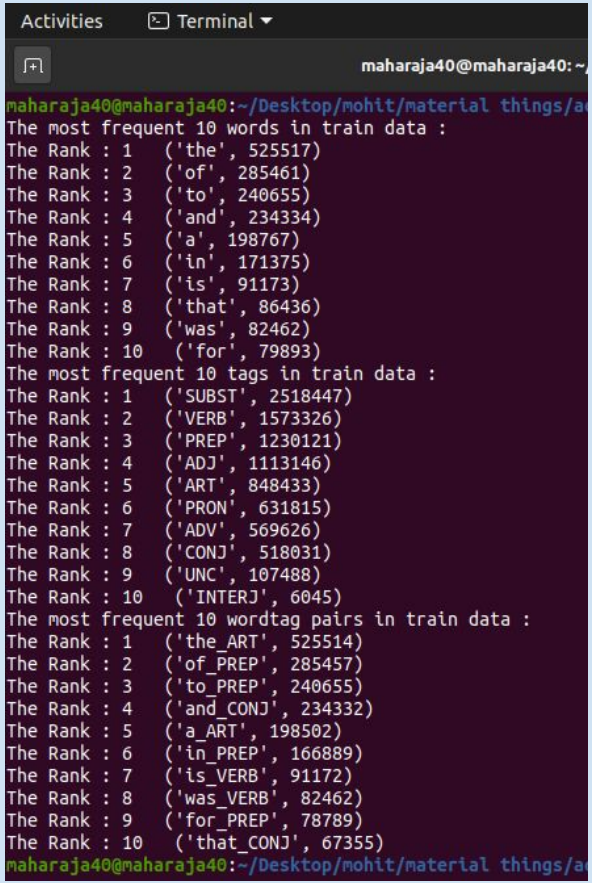


Figure depicts : Category wise Rankings

The most frequent 10 words in train data :
The Rank : 1 ('the', 525517)
The Rank : 2 ('of', 285461)
The Rank : 3 ('to', 240655)
The Rank : 4 ('and', 234334)
The Rank : 5 ('a', 198767)
The Rank : 6 ('in', 171375)
The Rank : 7 ('is', 91173)
The Rank : 8 ('that', 86436)
The Rank : 9 ('was', 82462)
The Rank : 10 ('for', 79893)

Mohit Kumar 18114049 | Sritam Behera 18116077 | OM Katiyar 18116057

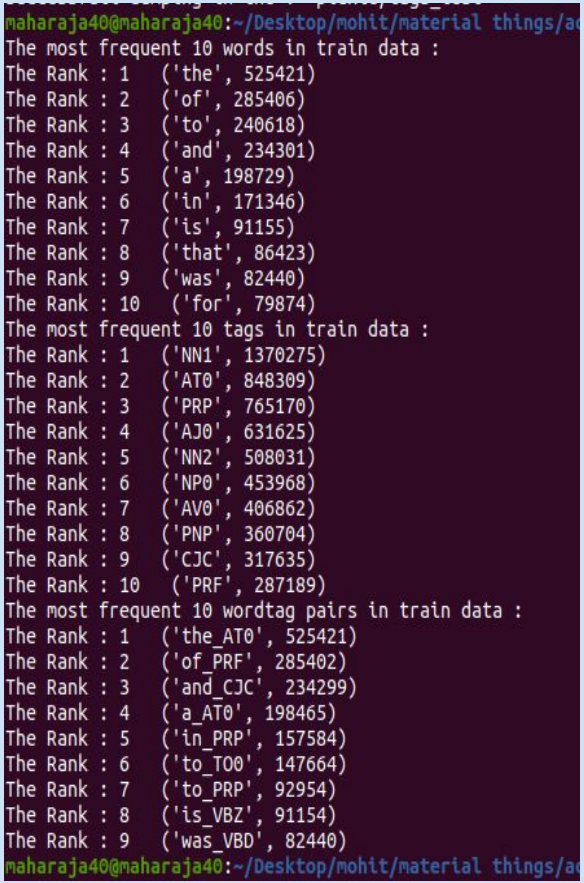


Figure depicts : Deeper Detailed Tag wise Rankings

The most frequent 10 Categories of POS in train data : (Category Wise)
The Rank : 1 ('SUBST', 2518447)
The Rank : 2 ('VERB', 1573326)
The Rank : 3 ('PREP', 1230121)
The Rank : 4 ('ADJ', 1113146)
The Rank : 5 ('ART', 848433)
The Rank : 6 ('PRON', 631815)
The Rank : 7 ('ADV', 569626)
The Rank : 8 ('CONJ', 518031)
The Rank : 9 ('UNC', 107488)
The Rank : 10 ('INTERJ', 6045)

The most frequent 10 tags in train data : (Deeper Tag Based Analysis)
The Rank : 1 ('NN1', 1370275)
The Rank : 2 ('AT0', 848309)
The Rank : 3 ('PRP', 765170)
The Rank : 4 ('AJ0', 631625)
The Rank : 5 ('NN2', 508031)
The Rank : 6 ('NP0', 453968)
The Rank : 7 ('AV0', 406862)
The Rank : 8 ('PNP', 360704)
The Rank : 9 ('CJC', 317635)
The Rank : 10 ('PRF', 287189)

A humble and small suggestion from our team :
We found this tag list more relevant as compared to provided tag list:
http://www.natcorp.ox.ac.uk/docs/gramtag.html
Provided tag list is correct no doubt, but this above has all tags that we found, whereas the given list lacks many.
We in no way, are denying your authority, As said this is humble suggestion.

Mohit Kumar 18114049 | Sritam Behera 18116077 | OM Katiyar 18116057

## The Analysis of Word & Tag Distribution :
## (Please refer the two images in previous slides)

1. Nouns Corresponds to the words who are in majority in the database.
2. But the Nouns are highly scattered, hence there is no Noun in the list of most common words in train data set.
3. The usage of Nouns is almost 1.6 times as compared to the Usage of Verbs.
4. Although database contains huge number of Verbs, just like Nouns they are highly scattered, this scattering for both (Nouns and Verbs) can become limitation for proper calculations of their pos tags due to shortage of data.
5. “Of” has been given a special tag “PRF”, because of its frequency and its almost exclusively post-nominal function.

Mohit Kumar 18114049 | Sritam Behera 18116077 | OM Katiyar 18116057