**Assignment 2**

Text analysis in R on non-English text

The Hindi document is based on a ’Vikram And Betal’ story.

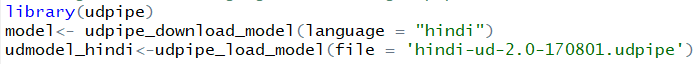
Note: The text files are in Hindi language. It should be placed in the same folder as that of ‘Assignment 2-Hindi.R’ file.

The assignment has been completed in the following steps:

Step 1: Import required R packages



Step 2: Loading Hindi language model using udpipe



Step 3: Load documents



**Analysis 1**: Find the frequency of each words and display the words with largest frequency

* Read the file using:



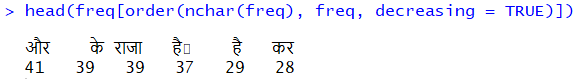
* Split the words using:



* Find the frequency of each word using:

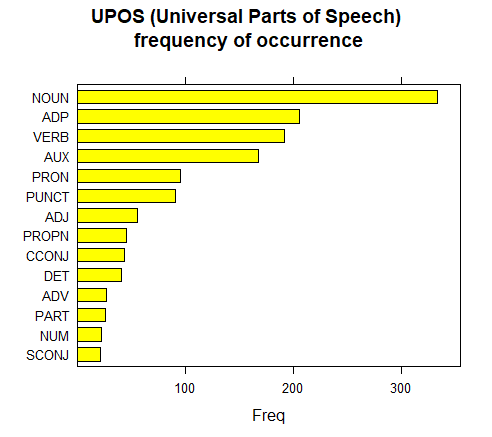


* Sort the words in descending order of frequency and display the top 7



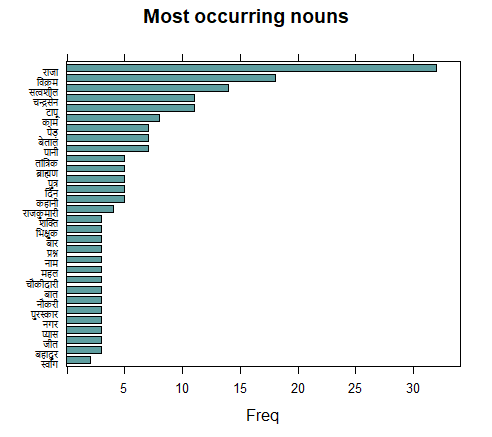
**Analysis 2**: Calculate the parts of speech

* Find the frequency of parts of speech and display it using a bar plot



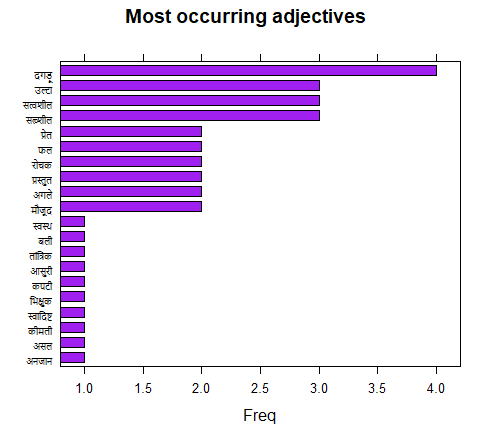
**Analysis 3**: Find the most occurring nouns

* Find the frequency of the nouns and display it using a bar plot

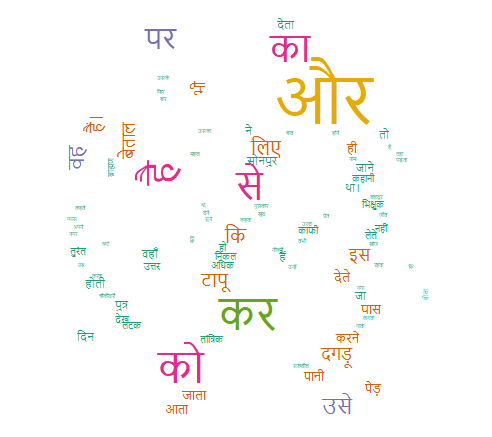


**Analysis 4**: Find the most occurring adjectives

* Find the frequency of the adjectives and display using a bar plot



**Analysis 5**: Display the most frequently occurring words with frequency greater than 3 in a WordCloud



**Analysis 6**: Find cooccurrences within a 3 word limit

