

Date: 30/11/21

Our main Problem Now:

↳ Part of Speech Tagging

↳ Read Multiple Paper:

↳ Rule base Technique (1)

↳ Bi-LSTM (2)

↳ Idea:

↳ If we get lexicon table of

Pashto language then just extract the tag dataset from table

↳ Problem:

↳ What we will do if our word is not present in lexicon table

↳ May be neglect that word

Algorithm:

↳

Algorithm 1 Algorithm for POS tagging

```

0: Read Input text from the file
0: Tokenize the text into tokens
0: Read each token in text.
0: for each token in lexicon do
1: if found then
2:   append the tag with a token
3: else if multiple tags exist then
4:   tag with given rules
5: end if
5: end for
5: Write the appended tag with a token to file.
5: If not ask for user to tag, otherwise leave blank
5: Add the tag with new word to the lexicon.

```

linked resources : <https://github.com/mohbadar/pashto-text-dataset>BBC Pashto : <https://www.bbc.com/pashto/afghanistan-59455535>research paper BiLSTM : <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9283773>

2222
2222

.

.