

✓ NLP_intro

```
import tensorflow as tf
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
```

```
sentences = [
    'i do not like dogs',
    'do you like dogs ?',
    'His cat is black',
    'This white cat is very cute',
    'gandhinagar गांधीनगर'
]
```

```
tokenizer = Tokenizer(oov_token='<oov>')
tokenizer.fit_on_texts(sentences)
word_index = tokenizer.word_index
print(word_index)
```

```
{'<oov>': 1, 'do': 2, 'like': 3, 'dogs': 4, 'cat': 5, 'is': 6, 'i': 7, 'not': 8, 'you': 9, 'his': 10, 'black': 11, 'this': 12, 'whit
```

```
sequences = tokenizer.texts_to_sequences(sentences)
print(sequences[0])
```

```
[7, 2, 8, 3, 4]
```

```
padded = pad_sequences(sequences, padding='post')
print(padded[0])
print(padded[1])
```

```
[7 2 8 3 4 0]
[2 9 3 4 0 0]
```