

Gensim

- It is a NLP package that does topic modelling.
- Given a sequence of words as a sentence, paragraph or a document, a dictionary of words can be created by:

```
import gensim
from gensim import corpora

documents = [list of strings]

texts = [[text for text in line.split()] for line in documents]

dictionary = corpora.Dictionary(texts)
```

- The word map of the dictionary can be seen by:

```
dictionary.token2id
```