```python
from IPython import get_ipython
from IPython.display import display
import gensim
from gensim.utils import simple_preprocess
from gensim.parsing.preprocessing import STOPWORDS
from nltk.stem import WordNetLemmatizer, SnowballStemmer
import nltk
nltk.download('wordnet')
nltk.download('punkt')
text_file = "%%writefile my_story.txt"
from google.colab import files
uploaded = files.upload()

def preprocess_text(text):
  tokens = gensim.utils.simple_preprocess(str(text), deacc=True)
  tokens = [token for token in tokens if token not in gensim.parsing.preprocessing.STOPWORDS]
  stemmer = SnowballStemmer("english")
  stemmed_tokens = [stemmer.stem(token) for token in tokens]
  lemmatizer = WordNetLemmatizer()
  lemmatized_tokens = [lemmatizer.lemmatize(token) for token in stemmed_tokens]
  return lemmatized_tokens

with open(text_file, 'r') as file:
  text_data = file.read()
preprocessed_text = preprocess_text(text_data)
print(preprocessed_text)
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]    Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]    Package punkt is already up-to-date!
  Choose Files  %%writefile my_story.txt
  • %%writefile my_story.txt(text/plain) - 67 bytes, last modified: 25/1/2025 - 100% done
  Saving %%writefile my_story.txt to %%writefile my_story (1).txt
  ['time', 'land', 'far', 'away', 'live', 'brave', 'knight']
```

Start coding or generate with AI.

Start coding or generate with AI.