

## Exp\_10\lemmatization.py

```
1 import nltk
2 from nltk.stem import WordNetLemmatizer
3 from nltk.corpus import wordnet
4
5 # Download required resources
6 nltk.download('wordnet', quiet=True)
7 nltk.download('omw-1.4', quiet=True)
8 nltk.download('averaged_perceptron_tagger', quiet=True)
9
10 def get_wordnet_pos(tag):
11     """Convert POS tag to wordnet format"""
12     if tag.startswith('J'):
13         return wordnet.ADJ
14     elif tag.startswith('V'):
15         return wordnet.VERB
16     elif tag.startswith('N'):
17         return wordnet.NOUN
18     elif tag.startswith('R'):
19         return wordnet.ADV
20     else:
21         return wordnet.NOUN
22
23 def lemmatize_text(text):
24     """Lemmatize text with POS tagging"""
25     lemmatizer = WordNetLemmatizer()
26     tokens = nltk.word_tokenize(text)
27     pos_tags = nltk.pos_tag(tokens)
28
29     lemmatized = [lemmatizer.lemmatize(word, get_wordnet_pos(pos)) for word, pos in
30 pos_tags]
31     return lemmatized
32
33 def simple_lemmatize(words):
34     """Simple lemmatization without POS tagging"""
35     lemmatizer = WordNetLemmatizer()
36     return [lemmatizer.lemmatize(word) for word in words]
37
38 if __name__ == "__main__":
39     sample_text = "The cats are running and jumping over the fences"
40
41     # With POS tagging
42     result = lemmatize_text(sample_text)
43     print("Lemmatized (with POS):", ' '.join(result))
44
45     # Without POS tagging
46     words = nltk.word_tokenize(sample_text)
47     simple_result = simple_lemmatize(words)
48     print("Lemmatized (simple):", ' '.join(simple_result))
```