

```
In [1]: !pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\pronn\anaconda3\lib\site-packages
(3.7)
Requirement already satisfied: click in c:\users\pronn\anaconda3\lib\site-packages
(from nltk) (8.0.4)
Requirement already satisfied: tqdm in c:\users\pronn\anaconda3\lib\site-packages
(from nltk) (4.64.1)
Requirement already satisfied: regex>=2021.8.3 in c:\users\pronn\anaconda3\lib\sit
e-packages (from nltk) (2022.7.9)
Requirement already satisfied: joblib in c:\users\pronn\anaconda3\lib\site-package
s (from nltk) (1.1.0)
Requirement already satisfied: colorama in c:\users\pronn\anaconda3\lib\site-packa
ges (from click->nltk) (0.4.5)
```

```
In [2]: import re
```

```
def porter_stemmer(word) :
    # Step 1a
    if word.endswith('sses') :
        word = re.sub('sses$', 'ss', word)
    elif word.endswith('ies') :
        word = re.sub('ies$', 'i', word)
    elif word.endswith('ss') :
        word = re.sub('ss$', 'ss', word)
    else :
        word = re.sub('s$', '', word)

    # Step 1b
    if re.search(r'[aeiouy].*ed$', word) :
        word = re.sub('ed$', '', word)
        if re.search(r'[aeiouy]', word) :
            word = re.sub('ing$', '', word)

    # Step 1c
    if re.search(r'[aeiouy].*y$', word) :
        word = re.sub('y$', 'i', word)

    # Step 2
    if re.search(r'[aeiouy].*(ational|tional|enci|anci|izer|bli|alli|entli|eli|ousl', word) :
        word = re.sub(r'(ational|tional|enci|anci|izer|bli|alli|entli|eli|ousli|ous', word)
        if re.search(r'[aeiouy].*(ational|tional|enci|anci|izer|bli|alli|entli|eli|ousl', word) :
            word = re.sub(r'(ational|tional|enci|anci|izer|bli|alli|entli|eli|ousli|ousl', word)

    # Step 3
    if re.search(r'[aeiouy].*(icate|ative|alize|iciti|ical|ful|ness)$', word) :
        word = re.sub(r'(icate|ative|alize|iciti|ical|ful|ness)$', '', word)
        if re.search(r'[aeiouy].*(icate|ative|alize|iciti|ical|ful|ness)$', word) :
            word = re.sub(r'(icate|ative|alize|iciti|ical|ful|ness)$', '', word)

    # Step 4
    if re.search(r'[aeiouy].*(al|ance|ence|er|ic|able|ible|ant|ement|ment|ent|ou|is', word) :
        word = re.sub(r'(al|ance|ence|er|ic|able|ible|ant|ement|ment|ent|ou|ism|ate', word)
        if re.search(r'[aeiouy].*(al|ance|ence|er|ic|able|ible|ant|ement|ment|ent|ou|is', word) :
            word = re.sub(r'(al|ance|ence|er|ic|able|ible|ant|ement|ment|ent|ou|ism|ate', word)
```

```

# Step 5a
if re.search(r'e$', word) :
    word = re.sub('e$', '', word)
    if re.search(r'[aeiouy].*e$', word) :
        word = re.sub('e$', '', word)

# Step 5b
if re.search(r'[aeiouy].*ll$', word) :
    word = re.sub('ll$', 'l', word)

return word

words_to_stem = ["walking", "swayed", "happily", "seen", "children"]
stemmed_words = [porter_stemmer(word) for word in words_to_stem]
print(stemmed_words)

```

```
['walking', 'swai', 'happili', 'seen', 'children']
```

```

In [3]: from nltk.stem import PorterStemmer

nltk_stemmer = PorterStemmer()

nltk_stemmed_words = [nltk_stemmer.stem(word) for word in words_to_stem]
print(nltk_stemmed_words)

```

C:\Users\pronn\anaconda3\lib\site-packages\scipy__init__.py:155: UserWarning: A NumPy version >=1.18.5 and <1.25.0 is required for this version of SciPy (detected version 1.26.0

```

    warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
['walk', 'sway', 'happili', 'seen', 'children']

```