- NLTK tokenization
  - link1
  - ```
    from nltk.tokenize import word_tokenize
    text = "God is Great! I won a lottery."
    print(word_tokenize(text))

    Output: ['God', 'is', 'Great', '!', 'I', 'won', 'a', 'lottery', '.']
    ```
- NLTK stop words removal
  - link1
  - ```
    from nltk.corpus import stopwords
    from nltk.tokenize import word_tokenize

    example_sent = "This is a sample sentence, showing off the stop words
    filtration."

    stop_words = set(stopwords.words('english'))

    word_tokens = word_tokenize(example_sent)

    filtered_sentence = [w for w in word_tokens if not w in stop_words]

    filtered_sentence = []

    for w in word_tokens:
        if w not in stop_words:
            filtered_sentence.append(w)

    print(word_tokens)
    print(filtered_sentence)

    # outputs
    ['This', 'is', 'a', 'sample', 'sentence', ',', 'showing',
    'off', 'the', 'stop', 'words', 'filtration', '.']
    ['This', 'sample', 'sentence', ',', 'showing', 'stop',
    'words', 'filtration', '.']
    ```
- NLTK stemming
  - link1
  - ```
    from nltk.stem import PorterStemmer
    from nltk.tokenize import word_tokenize
    ```
    ```

```
ps = PorterStemmer()

sentence = "Programers program with programing languages"
words = word_tokenize(sentence)

for w in words:
    print(w, " : ", ps.stem(w))

# outputs
Programers  :  program
program  :  program
with  :  with
programing  :  program
languages  :  languag
```

- Python learning
  - English version
    - an interactive learning website Check this
  - Chinese version
    - Check this
- Building inverted index
  - sample code