

Stopwords

```
In [44]: from nltk.corpus import stopwords
stop_words=set(stopwords.words("Chinese"))
print(stop_words)
```

{'共同', '起见', '多次', '表示', '某个', '哉', '她的', '倘然', '是不是', '除了', '这会儿', '即或', '这麼', '某', '省得', '每当', '那么', '最好', '比较', '造成', '一起', '可', '逐渐', '接着', '顺', '而况', '所以', '莫若', '越是', '经', '咳', '我们', '它的', '她', '本', '不变', '乃至', '而', '突出', '多少', '哪边', '其实', '矣', '宣布', '适应', '整个', '和', '哦', '大家', '一些', '怎麼', '你', '并且', '做到', '嘻', '由此可见', '适当', '据', '抑或', '能否', '至于', '她们', '为何', '维持', '哈哈', '出去', '咋', '有著', '其它', '今后', '而已', '帮助', '一时', '一则', '为', '冒', '有所', '这样', '强烈', '结合', '考虑', '俺', '过来', '随', '加之', '来', '出现', '经过', '离', '在下', '如下', '哟', '倘若', '朝着', '有关', '极了', '不久', '那边', '乃', '广大', '反过来说', '自从', '吗', '哩', '嘿', '给', '自身', '而外', '这点', '纵然', '沿着', '得出', '它们', '许多', '无宁', '起来', '似乎', '总的来说', '显然', '反过来', '有利', '甚至', '产生', '要么', '不光', '重新', '如', '则', '以上', '受到', '以为', '尽', '慢说', '这个', '乎', '巨大', '具体地说', '转变', '如其', '前者', '深入', '那时', '有些', '后来', '少数', '各个', '及其', '得到', '只要', '经常', '再说', '实现', '所', '不同', '我的', '岂但', '俺们', '知道', '冲', '坚决', '望', '促进', '那个', '强调', '为什麼', '完全', '何况', '注意', '後來', '再者', '连同', '应该', '一般', '及时', '呸', '其', '个别', '不怕', '最大', '与否', '不特', '满足', '么', '其次', '集中', '各人', '不足', '虽说', '因为', '心里', '既然', '被', '自各儿', '加强', '咚', '哪', '属于', '几时', '必要', '虽', '是的', '这么点儿', '于是乎', '向', '普通', '尚且', '进行', '最後', '宁可', '达到', '那', '别的', '代替', '各种', '有时', '召开', '良好', '那麼', '你们', '一旦', '似的', '总的来看', '另外', '下来', '啦', '几乎', '不如', '各级', '曾经', '不然', '运用', '进入', '进而', '防止', '不比', '喏', '鉴于', '其余', '呵', '先后', '高兴', '这种', '积极', '不单', '假如', '与此同时', '为了', '而言', '那儿', '周围', '哪儿', '要不是', '如何', '着呢', '不惟', '一致', '凭', '或', '各', '特殊', '看见', '真正', '别', '换句话说', '呃', '啊', '哈', '喔唷', '那么样', '上升', '倘或', '一来', '不够', '及', '哪些', '毋宁', '阿', '为着', '或者', '上来', '掌握', '没有', '随着', '上下', '巩固', '过', '万一', '要不', '者', '又', '呕', '有效', '具有', '认识', '咱', '看来', '如此', '往', '要', '连', '使得', '上述', '某些', '假使', '哗', '这么样', '谁知', '使用', '下面', '以后', '到', '就', '的', '任', '多数', '相当', '之一', '一面', '喂', '一下', '根本', '最近', '说明', '不管', '哎呀', '嘎', '漫说', '不一', '什么', '正在', '照', '先後', '故', '允许', '认真', '按', '总之', '不可', '个', '以', '这边', '应用', '只是', '更加', '乘', '们', '让', '除非', '明确', '扩大', '且', '吧', '而是', '移动', '实际', '紧接着', '即', '失去', '哼', '一片', '中间', '尔后', '与', '替', '相同', '真是', '着', '什麼', '另', '有力', '看看', '联系', '为什么', '只限', '诸位', '己', '一', '另一方面', '等', '那样', '需要', '比如', '以至', '加以', '相等', '准备', '部分', '只有', '形成', '何处', '之後', '从而', '待', '因此', '一次', '后面', '非但', '要求', '自家', '况且', '还有', '以免', '恰恰相反', '应当', '根据', '采取', '特别是', '继而', '个人', '必须', '以及', '安全', '乌乎', '然后', '嗯', '通常', '好象', '拿', '出来', '固然', '它们的', '当然', '处理', '由于', '愿意', '所有', '相应', '叮咚', '那些', '即若', '每天', '嗡嗡', '此外', '叫', '表明', '把', '其二', '呼哧', '倘使', '前面', '当时', '相似', '像', '及至', '接著', '一直', '怎', '不独', '腾', '不能', '然後', '人家', '啪达', '合理', '要不然', '显著', '较之', '当着', '果然', '相反', '嘘', '哼唷', '靠', '任何', '临', '嘎登', '仍然', '充分', '总结', '突然', '喂', '等等', '比方', '相对', '除', '这么', '然则', '按照', '打', '全部', '一边', '它', '一方面', '得', '换言之', '继续', '并不是', '一天', '向着', '看到', '谁', '彻底', '加入', '取得', '这', '保持', '绝对', '综上所述', '别说', '趁', '转贴', '主张', '一定', '嘛', '复杂', '至', '进步', '反之', '倘', '此时', '战斗', '若是', '不会', '彼此', '过去', '逐步', '他', '首先', '之前', '下去', '毫不', '主要', '既是', '行动', '兮', '大力', '既', '照着', '那里', '下列', '确定', '哪里', '避免', '任务', '各地', '却不', '果真', '咱们', '甚而', '不过', '对应', '唉', '顺着', '即便', '已经', '彼', '哪天', '哎', '依靠', '反应', '不

但', '几', '无法', '故此', '大约', '依', '边', '叫做', '什么样', '不成', '不是', '全面', '但', '左右', '赶', '管', '你的', '决定', '明显', '以下', '吓', '组成', '纵令', '立即', '哇', '觉得', '遇到', '规定', '大量', '要是', '非常', '了', '以后', '坚持', '不得', '总而言之', '上面', '专门', '开外', '第', '虽则', '企图', '鄙人', '甚么', '比', '欢迎', '何时', '怎样', '关于', '自个儿', '适用', '于是', '凭借', '常常', '范围', '能', '将', '若非', '一样', '后面', '设若', '有着', '目前', '罢了', '与其', '一切', '有', '每', '变成', '多', '每年', '吧哒', '方便', '方面', '任凭', '这些', '若', '跟', '认为', '争取', '可能', '纵', '也罢', '之类', '转动', '遭到', '吱', '看出', '从事', '互相', '用', '问题', '不要', '不问', '往往', '但是', '嗨', '丰富', '您', '里面', '具体', '十分', '当前', '宁愿', '相信', '这儿', '总是', '之后', '每个', '以前', '特点', '可见', '正常', '重要', '如若', '就是说', '如上所述', '噢', '原来', '同时', '该', '可是', '对于', '可以', '最高', '本着', '大批', '非徒', '并不', '必然', '怎么办', '由', '不只', '举行', '这就是说', '其他', '普遍', '先生', '其中', '不拘', '尽管', '了解', '大大', '呢', '宁肯', '时候', '构成', '之', '哪怕', '焉', '其一', '呜呼', '较', '今天', '朝', '纵使', '而且', '归', '地', '旁人', '即使', '各位', '以致', '今年', '他们', '他人', '虽然', '怎么', '啐', '作为', '现代', '以便', '基本', '最后', '这里', '同', '趁着', '啥', '意思', '哪年', '通过', '广泛', '于', '人们', '密切', '是', '容易', '是否', '清楚', '有的', '或是', '论', '宁', '两者', '起', '否则', '就是', '呀', '例如', '我', '即令', '那么些', '附近', '同一', '在', '如果', '分别', '上去', '以外', '以至于', '结果', '他的', '怎么样', '能够', '以来', '大多数', '并没有', '此间', '还是', '不仅', '哪个', '因', '并', '无论', '重大', '限制', '不断', '存在', '自己', '迅速', '同样', '开展', '沿', '各自', '为主', '之所以', '假若', '双方', '的话', '所谓', '哪样', '中小', '随著', '相对而言', '呜', '有点', '不论', '今後', '何', '成为', '此', '引起', '练习', '行为', '除此之外', '这时', '设使', '前后', '反映', '开始', '现在', '自', '正如', '好的', '云云', '前进', '严重', '从', '伟大', '尤其', '来着', '依照', '这么些', '那会儿', '严格', '也', '获得', '呗', '也是', '具体说来', '完成', '当', '因而', '直接', '不敢', '对', '也好', '然而', '直到', '总的说来', '哎哟'}

Removing Stopwords

```
In [18]: from nltk.tokenize import word_tokenize
text="""Hello Mr. Smith, how are you doing today? The weather is great, and city
The sky is pinkish-blue. You shouldn't eat cardboard"""
tokenized_sent=word_tokenize(text)

filtered_sent=[]
for w in tokenized_sent:
    if w not in stop_words:
        filtered_sent.append(w)
print("Tokenized Sentence:",tokenized_sent)
print("Filterd Sentence:",filtered_sent)
```

```
Tokenized Sentence: ['Hello', 'Mr.', 'Smith', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ',', 'and', 'city', 'is', 'awesome', '.', 'The', 'sky', 'is', 'pinkish-blue', '.', 'You', 'should', 'n't', 'eat', 'cardboard']
Filterd Sentence: ['Hello', 'Mr.', 'Smith', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky', 'pinkish-blue', '.', 'You', 'n't', 'eat', 'cardboard']
```

Stemming

```
In [19]: # Stemming
from nltk.stem import PorterStemmer
from nltk.tokenize import sent_tokenize, word_tokenize

ps = PorterStemmer()

stemmed_words=[]
for w in filtered_sent:
    stemmed_words.append(ps.stem(w))

print("Filtered Sentence:",filtered_sent)
print("Stemmed Sentence:",stemmed_words)
```

Filtered Sentence: ['Hello', 'Mr.', 'Smith', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky', 'pinkish-blue', '.', 'You', 'n't', 'eat', 'cardboard']
Stemmed Sentence: ['hello', 'mr.', 'smith', ',', 'today', '?', 'the', 'weather', 'great', ',', 'citi', 'awesom', '.', 'the', 'sky', 'pinkish-blu', '.', 'you', 'n't', 'eat', 'cardboard']

Lemmatization

```
In [35]: #Lexicon Normalization
#performing stemming and Lemmatization

from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()

from nltk.stem.porter import PorterStemmer
stem = PorterStemmer()

word = "Lemmatization"
print("Lemmatized Word:",lem.lemmatize(word,"v"))
print("Stemmed Word:",stem.stem(word))
```

Lemmatized Word: Lemmatization
Stemmed Word: lemmat

POS Tagging

```
In [21]: sent = "Albert Einstein was born in Ulm, Germany in 1879."
```

```
In [22]: tokens=nltk.word_tokenize(sent)
print(tokens)
```

```
['Albert', 'Einstein', 'was', 'born', 'in', 'Ulm', ',', 'Germany', 'in', '1879', '.']
```

```
In [23]: nltk.pos_tag(tokens)
```

```
Out[23]: [('Albert', 'NNP'),
          ('Einstein', 'NNP'),
          ('was', 'VBD'),
          ('born', 'VBN'),
          ('in', 'IN'),
          ('Ulm', 'NNP'),
          (',', ','),
          ('Germany', 'NNP'),
          ('in', 'IN'),
          ('1879', 'CD'),
          ('.', '.')]

```

NNP:Proper noun singular

VBD : Verb past tense

CD:Cardinal number

VBN:Verb past participle

```
In [ ]:
```