

```
In [9]: # Task 1 - define some dummy text, filter from stop words, make new filtered list, print new list
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import sent_tokenize, word_tokenize

EXAMPLE_TEXT = "This is yet another example of 3 dummy sentences. They make no sense whatsoever. But perhaps the results will yield something extraordinary. Who knows?"

# TEXT_IN_WORDS = word_tokenize(EXAMPLE_TEXT)
# print(TEXT_IN_WORDS)

stop_words = set(stopwords.words('english'))
word_tokens = word_tokenize(EXAMPLE_TEXT)
filtered_sentence3 = []
for w in word_tokens:
    if w not in stop_words:
        filtered_sentence3.append(w)

filtered_sentence3
print(filtered_sentence3)

['This', 'yet', 'another', 'example', '3', 'dummy', 'sentences', '.', 'They', 'make', 'sense', 'whatsoever', '.', 'But', 'perhaps', 'results', 'yield', 'so mething', 'extraordinary', '.', 'Who', 'knows', '?']
```

```
In [10]: # Task 1 - same but from loaded file "movie_reviews"
import nltk
from nltk.corpus import stopwords, movie_reviews
from nltk.tokenize import word_tokenize

text = movie_reviews.words("pos/cv000_29590.txt")

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

filtered_sentence4 = [w for w in text if not w in stop_words]

print(filtered_sentence4)
```

['films', 'adapted', 'comic', 'books', 'plenty', 'success', ',', ',', 'whether',
 "", 'superheroes', '(', 'batman', ',', ',', 'superman', ',', ',', 'spawn', ')', ',', ',',
 'geared', 'toward', 'kids', '(', 'casper', ')', 'arthouse', 'crowd', '(', 'gh
 ost', 'world', ')', ',', ',', "", 'never', 'really', 'comic', 'book', 'like', 'he
 ll', '.', 'starters', ',', ',', 'created', 'alan', 'moore', '(', 'eddie', 'campbel
 l', ')', ',', ',', 'brought', 'medium', 'whole', 'new', 'level', 'mid', "", '80
 s', '12', '-', 'part', 'series', 'called', 'watchmen', '.', 'say', 'moore',
 'campbell', 'thoroughly', 'researched', 'subject', 'jack', 'ripper', 'would',
 'like', 'saying', 'michael', 'jackson', 'starting', 'look', 'little', 'odd',
 '.', 'book', '(', 'graphical', 'novel', ',', ',', "", ')', '500', 'pages', 'lo
 ng', 'includes', 'nearly', '30', 'consist', 'nothing', 'footnotes', '.', 'wor
 ds', ',', ',', "", 'dismiss', 'film', 'source', '.', 'get', 'past', 'whole', 'com
 ic', 'book', 'thing', ',', ',', 'might', 'find', 'another', 'stumbling', 'block',
 'hell', "", 'directors', ',', ',', 'albert', 'allen', 'hughes', '.', 'getting',
 'hughes', 'brothers', 'direct', 'seems', 'almost', 'ludicrous', 'casting', 'c
 arrot', 'top', ',', ',', 'well', ',', ',', 'anything', ',', ',', 'riddle', ':', 'better', 'd
 irect', 'film', "", 'set', 'ghetto', 'features', 'really', 'violent', 'stree
 t', 'crime', 'mad', 'geniuses', 'behind', 'menace', 'ii', 'society', '?', 'gh
 etto', 'question', ',', ',', 'course', ',', ',', 'whitechapel', '1888', 'london', "",
 'east', 'end', '.', 'filthy', ',', ',', 'sooty', 'place', 'whores', '(', 'cal
 led', "", 'unfortunates', "", ')', 'starting', 'get', 'little', 'nervous',
 'mysterious', 'psychopath', 'carving', 'profession', 'surgical', 'precision',
 '.', 'first', 'stiff', 'turns', ',', ',', 'copper', 'peter', 'godley', '(', 'robbi
 e', 'coltrane', ',', ',', 'world', 'enough', ')', 'calls', 'inspector', 'frederic
 k', 'abberline', '(', 'johnny', 'depp', ',', ',', 'blow', ')', 'crack', 'case',
 '.', 'abberline', ',', ',', 'widower', ',', ',', 'prophetic', 'dreams', 'unsuccessfull
 y', 'tries', 'quell', 'copious', 'amounts', 'absinthe', 'opium', '.', 'upon',
 'arriving', 'whitechapel', ',', ',', 'befriends', 'unfortunate', 'named', 'mary',
 'kelly', '(', 'heather', 'graham', ',', ',', 'say', "", ')', 'proceeds', 'investi
 gate', 'horribly', 'gruesome', 'crimes', 'even', 'police', 'surgeon', "", 's
 tomach', '.', 'think', 'anyone', 'needs', 'briefed', 'jack', 'ripper',
 ',', ',', "", 'go', 'particulars', ',', ',', 'say', 'moore', 'campbell', 'unique', 'in
 teresting', 'theory', 'identity', 'killer', 'reasons', 'chooses', 'slay',
 '.', 'comic', ',', ',', "", 'bother', 'cloaking', 'identity', 'ripper', ',', ',', 'scr
 eenwriters', 'terry', 'hayes', '(', 'vertical', 'limit', ')', 'rafael', 'ygle
 sias', '(', 'les', 'mis', '?', 'rables', ')', 'good', 'job', 'keeping', 'hidd
 en', 'viewers', 'end', '.', 'funny', 'watch', 'locals', 'blindly', 'poin
 t', 'finger', 'blame', 'jews', 'indians', ',', ',', ',', 'englishman', 'could', 'n
 ever', 'capable', 'committing', 'ghastly', 'acts', '.', 'hell', "", 'endin
 g', 'whistling', 'stonecutters', 'song', 'simpsons', 'days', '(', 'hold
 s', 'back', 'electric', 'car', '/', 'made', 'steve', 'gutenberg', 'star',
 '?', 'worry', '-', 'make', 'sense', 'see', '.', 'ont
 o', 'hell', 'appearance', ':', 'certainly', 'dark', 'bleak', 'enoug
 h', ',', ',', 'surprising', 'see', 'much', 'looks', 'like', 'tim', 'burton',
 'film', 'planet', 'apes', '(', 'times', ',', ',', 'seems', 'like', 'sleepy', 'holl
 ow', '2', ')', '.', 'print', 'saw', 'completely', 'finished', '(', 'colo
 r', 'music', 'finalized', ',', ',', 'comments', 'marilyn', 'manson', ')', ',', ',', 'ci
 nematographer', 'peter', 'deming', '(', 'say', 'word', ')', 'ably', 'cap
 tures', 'dreariness', 'victorian', '-', 'era', 'london', 'helped', 'make', 'f
 lashy', 'killing', 'scenes', 'remind', 'crazy', 'flashbacks', 'twin', 'peak
 s', ',', ',', 'even', 'though', 'violence', 'film', 'pales', 'comparison', 'blac
 k', '-', '-', 'white', 'comic', '.', 'oscar', 'winner', 'martin', 'childs',
 "", '(', 'shakespeare', 'love', ')', 'production', 'design', 'turns', 'origi
 nal', 'prague', 'surroundings', 'one', 'creepy', 'place', '.', 'even', 'actin
 g', 'hell', 'solid', ',', ',', 'dreamy', 'depp', 'turning', 'typically', 'strong',
 'performance', 'deftly', 'handling', 'british', 'accent', '.', 'ians', 'hol
 m', '(', 'joe', 'gould', 'secret', ')', 'richardson', '(', '102', 'dalma

```
tians', ')', 'log', 'great', 'supporting', 'roles', ',,', 'big', 'surprise',  
'graham', '.', 'cringed', 'first', 'time', 'opened', 'mouth', ',,', 'imaginin  
g', 'attempt', 'irish', 'accent', ',,', 'actually', '"', 'half', 'bad', '.',  
'film', ',,', 'however', ',,', 'good', '.', '2', ':', '00', '-', 'r', 'strong',  
'violence', '/', 'gore', ',,', 'sexuality', ',,', 'language', 'drug', 'conten  
t']
```