

```

import nltk
from collections import defaultdict
# Assuming we're working with English
from nltk.stem.snowball import EnglishStemmer
class Index:
    """ Inverted index datastructure """

    def __init__(self, tokenizer, stemmer=None, stopwords=None):
        """
        tokenizer    -- NLTK compatible tokenizer function
        stemmer      -- NLTK compatible stemmer
        stopwords    -- list of ignored words
        """
        self.tokenizer = tokenizer
        self.stemmer = stemmer
        self.index = defaultdict(list)
        self.documents = {}
        self.__unique_id = 0
        if not stopwords:
            self.stopwords = set()
        else:
            self.stopwords = set(stopwords)

    def lookup(self, word):
        """
        Lookup a word in the index
        """
        word = word.lower()
        if self.stemmer:
            word = self.stemmer.stem(word)

        return [self.documents.get(id, None) for id in
self.index.get(word)]

    def add(self, document):
        """
        Add a document string to the index
        """
        for token in [t.lower() for t in nltk.word_tokenize(document)]:
            if token in self.stopwords:
                continue

            if self.stemmer:
                token = self.stemmer.stem(token)

            if self.__unique_id not in self.index[token]:
                self.index[token].append(self.__unique_id)

```

```

        self.documents[self.__unique_id] = document
        self.__unique_id += 1

index = Index(nltk.word_tokenize,
              EnglishStemmer(),
              nltk.corpus.stopwords.words('english'))
# TOP10 Dire straits
index.add('Industrial Disease')
index.add('Private Investigations')
index.add('So Far Away')
index.add('Twisting by the Pool')
index.add('Skateaway')
index.add('Walk of Life')
index.add('Romeo and Juliet')
index.add('Tunnel of Love')
index.add('Money for Nothing')
index.add('Sultans of Swing')

# TOP10 Led Zeppelin
index.add('Stairway To Heaven')
index.add('Kashmir')
index.add('Achilles Last Stand')
index.add('Whole Lotta Love')
index.add('Immigrant Song')
index.add('Black Dog')
index.add('When The Levee Breaks')
index.add('Since I\'ve Been Lovin\' You')
index.add('Since I\'ve Been Loving You')
index.add('Over the Hills and Far Away')
index.add('Dazed and Confused')

# Let's make some queries:
print(index.lookup('song'))
# ['Immigrant Song']

print(index.lookup('daze'))
# ['Dazed and Confused']

print(index.lookup('confusion'))
# ['Dazed and Confused']

```

Output:

```

[1]: runfile('C:/Users/Admin/Desktop/swarupa/exp4.py',
wdir='C:/Users/Admin/Desktop/swarupa')

```

```
C:\ProgramData\Anaconda3\lib\site-packages\scipy\__init__.py: 138:  
UserWarning: A NumPy version 1.16.5 and <1.23.0 is required for this  
version of SciPy (detected version 1.23.1) warnings.warn(f"A NumPy  
version >={np_minversion) and <{np_maxversion) is required for this  
version of "
```

```
['Immigrant Song']
```

```
['Dazed and Confused']
```

```
['Dazed and Confused']
```