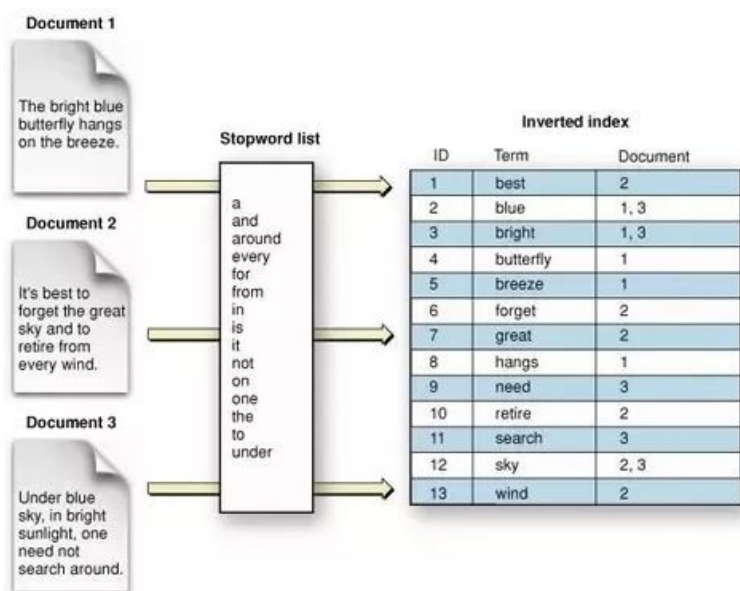


ایندکس معکوس (Inverted Index)



سید محمد حسین هاشمی ۴۰۲۲۳۶۳۱۴۳

فروردین ۱۴۰۳

فهرست مطالب

۲	۱	کلاس Inverted Index
۲	۲	متد <code>—init—</code>
۳	۳	متد <code>read-from-file</code>
۴	۴	متد <code>search</code>
۵	۵	خروجی

۱ کلاس Inverted Index

```
inverted_index.py

1 class InvertedIndex:
2     """ A very simple inverted index. """
```

تمام کدهای مربوطه در این کلاس نوشته می‌شود

۲ متد __init__

```
inverted_index.py

1 def __init__(self):
2     """ Create an empty inverted index. """
3
4     self.invertedIndex = {}
```

در اینجا یک دیکشنری خالی برای ذخیره ایندکس‌های معکوس ساخته می‌شود.

۳ متد read-from-file

```
inverted_index.py

1 def read_from_txt(self, file_name):
2     """ Construct index from given file
3
4     >>> ii = InvertedIndex()
5     >>> ii.read_from_txt('Datasets/example.txt')
6     >>> ii.invertedIndex
7     {'first': {1}, 'document': {1, 2, 3}, 'second': {2}, 'third': {3}}
8     """
9     with open(file_name, encoding="utf8") as file:
10         record_id = 0
11         for line in file:
12             record_id += 1
13             for word in re.split('[^a-zA-z]', line):
14                 if len(word) > 0:
15                     word = word.lower()
16                     if word not in self.invertedIndex:
17                         self.invertedIndex[word] = set()
18                     self.invertedIndex[word].add(record_id)
```

در این متد فایل حاوی متون دریافت و خوانده می‌شود و لازم به ذکر است که تمامی متحوا درون یک خط قرار می‌گیرد.

۴ متد search

```
inverted_index.py

1 def search(self, search):
2     """ Search with inverted indexes
3
4     >>> ii = InvertedIndex()
5     >>> ii.read_from_txt('Datasets/example.txt')
6     >>> ii.search('first')
7     {1}
8     """
9     search = map(lambda x: x.lower(), re.split('[^a-zA-z]', search))
10
11     results = set()
12     available_keys = self.invertedIndex.keys()
13     for key in search:
14         if key in available_keys:
15             if len(results) == 0:
16                 results = set(x for x in self.invertedIndex[key])
17             else:
18                 results.intersection_update(self.invertedIndex[key])
19     return results
```

در این متد عملیات اشتراک بین posting list ها انجام می‌شود و نتیجه جست‌وجو برگشت داده می‌شود.

۵ خروجی

```
inverted_index.py

1 if __name__ == '__main__':
2     if len(sys.argv) != 2:
3         print("Usage: python inverted_index.py [file_name]")
4         sys.exit(1)
5
6     filename = sys.argv[1]
7
8     ii = InvertedIndex()
9     ii.read_from_txt(filename)
10
11     status = input('See inverted index [0] or search [1]: ')
12     while status.isdigit() and (int(status) == 0 or int(status) == 1):
13         if status == '0':
14             print('inverted index')
15             print('input'.ljust(30), 'repeats'.ljust(10), 'documents')
16             for word, indexes in ii.invertedIndex.items():
17                 print(word.ljust(30), str(len(indexes)).ljust(10), indexes)
18         else:
19             for doc in ii.search(input('search: ')):
20                 print(doc)
21
22     status = input('See inverted index [0] or search [1] or any key to exit: ')
23     print('Bye')
```

همانطور که مشاهده می‌شود ترمینال ساده ای جهت کارکردن با برنامه ایجاد شده است.