optimizing_code_common_books

November 29, 2019

1 Optimizing Code: Common Books

Here's the code your coworker wrote to find the common book ids in `books_published_last_two_years.txt` and `all_coding_books.txt` to obtain a list of recent coding books.

```
In [2]: import time
        import pandas as pd
        import numpy as np

In [3]: alist = [[4],[2,4,6,8,10]]
        np_arr = np.array(alist)
        np_arr[1]=[5,6]
        np_arr

Out[3]: array([[4], [5, 6]], dtype=object)

In [4]: with open('books_published_last_two_years.txt') as f:
            recent_books = f.read().split('\n')

        with open('all_coding_books.txt') as f:
            coding_books = f.read().split('\n')

In [5]: start = time.time()
        recent_coding_books = []

        for book in recent_books:
            if book in coding_books:
                recent_coding_books.append(book)

        print(len(recent_coding_books))
        print('Duration: {} seconds'.format(time.time() - start))
```

```
96
Duration: 16.74492120742798 seconds
```

1.0.1 Tip #1: Use vector operations over loops when possible

Use numpy's `intersect1d` method to get the intersection of the `recent_books` and `coding_books` arrays.

```
In [7]: start = time.time()
        recent_coding_books = [] # TODO: compute intersection of lists
        # convert the list recent_books to numpy array
        r_arr = np.array(recent_books)
        c_arr = np.array(coding_books)
        rc_arr = np.intersect1d(r_arr,c_arr)
        recent_coding_books = rc_arr.tolist()
        print(len(recent_coding_books))
        print('Duration: {} seconds'.format(time.time() - start))
```

```
96
Duration: 0.03515338897705078 seconds
```

1.0.2 Tip #2: Know your data structures and which methods are faster

Use the set's `intersection` method to get the common elements in `recent_books` and `coding_books`.

```
In [8]: start = time.time()
        recent_coding_books = [] # TODO: compute intersection of lists
        # convert the list recent_books to set
        r_set = set(recent_books)
        c_set = set(coding_books)
        rc_set = c_set.intersection(r_set)
        recent_coding_books = list(rc_set)
        print(len(recent_coding_books))
        print('Duration: {} seconds'.format(time.time() - start))
```

```
96
Duration: 0.014719963073730469 seconds
```

```
In [ ]:
```