

# 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 [1]: import time
import pandas as pd
import numpy as np
```

```
In [2]: 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 [3]: 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: 16.75550389289856 seconds

## 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 [4]: start = time.time()
recent_coding_books = np.intersect1d(recent_books, coding_books) # TODO: compute
print(len(recent_coding_books))
print('Duration: {} seconds'.format(time.time() - start))
```

96

Duration: 0.03733658790588379 seconds

## 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 [5]: start = time.time()
recent_coding_books = list(set(recent_books).intersection(coding_books))# TODO:
print(len(recent_coding_books))
print('Duration: {} seconds'.format(time.time() - start))
```

96

Duration: 0.011070489883422852 seconds

In [ ]: