

تکلیف عملی اول

رسول صالحی

مجتبی ملائی

۱۴۰۱۳۱۳۸۳

۱ سوال اول

برای اینکه فایل متفاوت را بدست آوریم میتوانیم از hash استفاده کنیم. زیرا هر دو فایل یکسان هش یکسانی دارند ولی دو فایل متفاوت با احتمال بسیار زیادی هش متفاوتی خواهند داشت. hash ها به دلیل اندازه ثابت و کوچکشان برای مقایسه راحت تر هستند. از کد زیر برای پیدا کردن فایل مورد نظر استفاده شده است.

```
1 import hashlib
2
3 def hash_string(string): # Generate a hash for a string
4     return hashlib.md5(string.encode()).hexdigest()
5
6 file_path_prefix = './Q1/file'
7 file_count = 5000
8 files = {}
9
10 for i in range(1,file_count+1):
11     file_path = file_path_prefix + str(i)
12     with open(file_path, 'r') as file:
13         hash = hash_string(file.read()) # Generate hash for the file content
14         if hash in files : # If the hash already exists, append the file path to the list
15             files[hash].append(file_path)
16         else :
17             files[hash] = [file_path] # If the hash does not exist, create a new list with the file path
18
19 for key in files: # Iterate through the dictionary
20     if len(files[key]) == 1: # If the list has only one file, it is unique
21         different_file = files[key][0]
22         print(f'The different file is {different_file}')
23         with open (different_file, 'r') as file:
24             print(file.read())
```

Listing 1: Search for the unique file

خروجی کد:

```
1 The different file is ./Q1/file2745
2 empty
```

در کد بالا از md5 به دلیل سرعت بالاتر استفاده شده است.