ASSIGNMENT REPORT 2

Oğuzhan Akdoğan

oguzhanakdogan2@mu.edu.tr

Sunday 7thJune, 2020

1. INTRODUCTİON

This assignmentwhich the given us, we see how to understand process running in a program and multiprocessing implementation and also some process situation(orphan process).

2. ASSIGNMENTS

In this section, i will show what i do in this assignment with articles.

2.1 Child process and parent process

To creating child process, we have to use fork function of (os.fork())
To check which proccess we are in, just control the pid number;
İf pid is equal to 0, we are in child process, if greater than 0, then we are in parent process. Also this creates another process which will resume at exactly the same place as this one so, as if you call this method twice. How do we understand? Just call the method and control both pid == 0 and pid >0, program will execute these two control block.

2.2 Download Files via URL in array

To download a list of URL, we have to initialize an array to keep the URL inside. Then call each index of the array to download with "request.get(url,)" command. But this command is in the child process.

2.3 Avoid Orphan Process

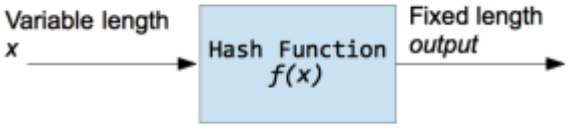
Orphan process is the situation when the parent finishes before child.If we encounter this process, the program execution in child process can not be completed. To avoid this kind of situation, we simply have to wait after killing the parent process using "**os.wait()**" command.

2.4 Control duplicate files

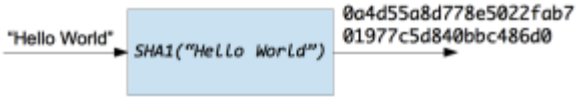
This program includes codes downloading some files by using their URL's in array. But some files are the same with other files. This will cause downloading some files twice. We have to handle this process by using "hashlib".

2.4.1 What is hashlib in python?

A hash function is a function that takes input of a variable length sequence of bytes and converts it to a fixed length sequence. This function something like "**meat grinder**".



Now, suppose that we have a string variable(which is "hello world") and we want to hash this variable by using **SHA1 Function.**



"Note that whichever variable you want to hash with SHA1, this function convert it into FIXED length sequence. This process is in force for all other hash functions."

2.4.2 What is "hashlib" used for?

Hash functions are used inside some cryptographic algorithms, in digital signatures, message authentication codes, manipulation detection, fingerprints, checksums (message integrity check), hash tables, password storage and much more. As a Python programmer we need this function to check for duplicate data.

We will use MD5 hash function.

MD5: Message digest algorithm producing a 128 bit hash value. This is widely used to check data integrity.

CONCLUSION

In conclusion i used os, multi processing, requests and haslib library in python.

REFERENCES

https://www.pythoncentral.io/hashing-strings-with-python/
https://pages.mtu.edu/~shene/FORUM/Taiwan-Forum/ComputerScience/003-OS/NOTES/process/fork/create.html

https://www.geeksforgeeks.org/creating-child-process-using-fork-python/