요구사항 명세서

TEDDY Kim

11/19/2021(금)

**프로젝트 - 디렉토리 내 바이너리 데이터 비교**

조건

* Linux 환경 & python 언어 활용 & SHA256 해시 사용
* 디렉토리는 하위 디렉토리가 없는 폴더개념
* 디렉토리 내 파일들은 전부 바이너리 파일 exe
* 동일 컨텐츠를 가지고 있는 파일은 같은 해시값을 가짐

요구사항

1. Hashing을 이용한 Hash value 비교
2. 해시값이 다르면 다른 파일(고정)
3. 해시값이 같으면

같은 파일 일수 있어서 파일명을 가지고 한 번 더 비교하여 **출력**

1. 11/24/2021(수) 데이터 테스트 예정

의견

* 기본적으로 hash sum을 구하기 위해서는 파일의 전체 내용을 읽어야 한다. 하지만 우리의 목적은 위 tuple을 “정확하게” 구하는 것이 아니라 중복 파일인지 여부를 판단하는 것이므로 hash sum을 모두 계산할 이유는 없다.
* 쉽게 파일 사이즈가 다르면 다른 파일로 간주할 수 있고, 처음 chunk (크기는 적당히 1MB)의 hash만 봐도 되는 것이다.