

중간고사

- 문제 : homework2.txt 파일을 이용하여 다음과 같은 조건을 만족하는 프로그램을 작성하시오.

- 조건 1 : 파일명(절대패스, 예) "c:\temp\output2.csv"을 입력으로 받아 모든 식품명의 월별 출현 빈도를 csv형식으로 저장 (첫줄은 헤더로 grocery_name, 월로 표시. 식품명은 중복을 허용하지 않으며, 모든 식품명을 대상으로 함) (7점)

```
grocery_name,1,2,3,4,5,6,7,8,9,10,11,12
grocery_name1,3,0,3,2,3,3,11,2,3,0,0,12
grocery_name2,3,0,0,2,3,3,0,2,3,0,0,12
grocery_name3,0,0,3,11,3,3,11,2,3,12,0,0
```

- 조건 2 : 2개의 식품명을 입력으로 받아, 월별로 같이 판매된 빈도를 계산하여 리스트로 반환 (리스트의 크기는 12이며, 1월부터 12월 순으로 저장) (7점)
 - 같이 판매된 기준은 같은 라인에 두 개의 식품명이 같이 사용된 경우로 함
- 조건 3 : 식품명을 입력으로 받아, 월별로 판매된 빈도를 계산하여 리스트로 반환 (리스트의 크기는 12이며, 1월부터 12월 순으로 저장) (5점)
- 조건 4 : 두 식품간의 유사도(Similarity)를 다음과 같은 수식(**$J(S,T)$ 계산 식** 이용)으로 계산 (식품명 두개를 입력으로 받아, 실수 형식의 유사도 값을 반환) (6점)

Weighted Jaccard distance between two vectors with non-negative entries is defined as $1 - J(S, T)$ where J is the Jaccard Similarity

$$J(S, T) = \frac{\sum_k \min(S_k, T_k)}{\sum_k \max(S_k, T_k)} \quad (1)$$

중간고사

- **조건 5** : 모든 식료품의 조합(combination)을 대상으로 유사도를 계산한 후, 입력 파라미터로 제공되는 한계값(threshold) 보다 큰 식품명의 조합을 다음과 같이 출력 (5점)

```
식품명1, 식품명2, 0.6242  
식품명1, 식품명3, 0.5322  
식품명1, 식품명5, 0.5224  
...
```

- **조건 6** : 파일의 작성시 클래스 개념을 이용하라. 한 개 이상의 클래스를 작성한 후, 정의된 클래스를 이용하여 조건1부터 조건5까지의 실행 결과를 확인할 수 있도록 적절한 메서드를 호출한 후 결과를 출력하는 메인 프로그램을 작성하라 (4점)
- **조건 7** : 컴프리헨션을 적절히 사용하여, 코드를 최대한 단순화 시켜라. 또한 적절한 데이터 타입을 선택하여 불필요한 데이터 변환 및 처리를 없애라. (2점)
- **조건 8** : 프로그램의 구조와 실행 방식에 대한 설명을 코드 내 주석 추가와 문서(2페이지 이내)로 작성하라 (4점)