

# Java Collection Framework

## 1. 숙제 5.1: 문제 설명

- 영어 문장으로 구성된 파일을 읽어서 가장 많이 등장하는 k-개의 단어와 빈도수를 출력
  - JFileChooser()를 이용하여 입력 파일 선택
  - 입력 1: 최소 문자수(minlen)  $\leftarrow$  minlen보다 짧은 단어는 고려하지 않음
  - 입력 2: 정수 k
- 출력: “길이  $\geq$  minlen”인 단어 중에서 가장 많이 나타나는 k-개의 단어들에 대해 빈도수의 내림차순으로 (단어, 빈도수)의 쌍을 출력

## 2. 숙제 5.2: 문제 설명

- 장바구니를 분석하여 가장 많이 구매하는 물건의 쌍을 k-개 조사
- 입력 파일의 내용
  - 정수 n: 장바구니의 수
  - (장바구니 id, 구매 물건 수 m, m 개의 물건 id 리스트)
    - 예: 100 3 27 96 1  $\leftarrow$  100번 장바구니에 1, 27, 96의 물건의 포함
  - $1 \leq \text{물건 id} \leq 10000$ ,  $1 \leq \text{장바구니 id} \leq 100000$
- 입력
  - JFileChooser()를 이용하여 입력 파일 선택
  - 입력 1: 정수 k
  - 입력 2: 신뢰도(confidence)  $\leftarrow$  실수 ( $0.0 \leq \text{신뢰도} \leq 1.0$ )
  - 입력 3: 지지도(support)  $\leftarrow$  실수 ( $0.0 \leq \text{신뢰도} \leq 1.0$ )
- 출력
  - 출력 1: 가장 많이 구매하는 k-개의 물건의 쌍과 빈도수를 출력
  - 출력 2: 지지도와 신뢰도를 만족하는 상관관계 리스트
    - 상관관계:  $10 \rightarrow 20$  (10번을 산 사람은, 20번도 높은 확률로 사더라...)
    - 지지도: 10과 20을 모두 포함한 장바구니의 비율
    - 신뢰도: 10이 포함된 장바구니중에서 20이 포함된 장바구니의 비율

## 2. 제출 내용

- 두 개의 소스 파일(Hw5\_1.java, Hw5\_2.java)을 Hw5.zip으로 묶어 제출

## 3. 평가: 140점 만점

- 숙제 5.1 (40점 만점)

- 결과값이 정확하게 출력: 40점
- 감점 대상:
  - Java Collection Framework의 Class들을 사용하지 않거나, 비효율적으로 사용
  - minlen과 k의 동작 오류
  - 소스 파일 이름 오류
  - 컴파일 오류 발생 등

- 숙제 5.2 (100점 만점)

- 출력 1의 결과값이 정확하게 출력: 40점
- 출력 2의 결과값이 정확하게 출력: 60점
- 감점 대상:
  - Java Collection Framework의 Class들을 사용하지 않거나, 비효율적으로 사용
  - k, 지지도, 신뢰도 등의 동작 오류
  - 소스 파일 이름 오류
  - 컴파일 오류 발생 등

- 별도로 package를 만들지 말고, default package를 이용하여 프로그램을 작성할 것.