

## [숙제02] Linux 및 AWK 연습

언어데이터과학

2023년 10월 16일 월요일 13시까지

- 다음의 두 파일을 작성하여 자신의 LDS2023 레포지토리에 업로드하라.
  - hw/hw02/hw02-answers.sh
  - hw/hw02/print-recipes.awk
- 위의 두 파일의 원본은 아래 링크의 폴더에서 다운로드할 수 있다.
  - <https://github.com/suparklingmin/LingDataSci2023/tree/main/hw/hw02>

### 0 데이터

4강 실습에서 만든 data/apicius/recipes/ 폴더에 포함된 모든 텍스트 파일(\*.txt)을 대상으로 한다.

#### 1 hw02-answers.sh 파일 6행 수정 (1점)

grep을 사용하여 레시피에 “GARUM”이 들어간 요리의 개수를 출력하는 명령을 작성하라. bash shell에서 실행했을 때 아래와 같이 8이 출력되면 된다.

```
$ grep # COMPLETE THIS CODE
8
```

#### 2 hw02-answers.sh 파일 9행 수정 (1점)

egrep을 사용하여 제목에 “FISH”가 들어간 모든 요리의 레시피에서 가장 자주 나온 단어 10개를 사용 빈도와 함께 출력하는 명령을 작성하라. bash shell에서 실행했을 때 아래와 같은 내용이 출력되면 된다.

```
$ egrep # COMPLETE THIS CODE
89 AND
42 WITH
40 THE
40 PEPPER
31 BROTH
30 OIL
27 WINE
26 FISH
25 VINEGAR
24 LOVAGE
```

#### 3 hw02-answers.sh 파일 12행 수정 (2점)

egrep을 사용하여 레시피 전체에서 트라이그램(trigram, 단어 세 개의 연쇄)의 출현 빈도를 세고, 빈도의 내림차순으로 정렬한 결과를 hw/hw02/apicius-3grams.txt 파일로 저장하는 명령을 작성하라. bash shell에서 실행했을 때 아래와 같은 내용이 출력되면 된다.

```
$ egrep # COMPLETE THIS LINE
$ head -3 hw/hw02/apicius-3grams.txt
57 SPRINKLE WITH PEPPER
29 AND A LITTLE
25 BROTH AND OIL
```

#### 4 print-recipes.awk 파일 완성 (1점)

data/apicius/recipes.txt 파일의 내용이 아래의 형식으로 출력될 수 있도록 print-recipes.awk 파일의 AWK 스크립트를 수정하라.

```
[요리 번호] 영어 요리 이름  
라틴어 요리 이름  
레시피  
  
[요리 번호] 영어 요리 이름  
라틴어 요리 이름  
레시피  
  
[요리 번호] 영어 요리 이름  
라틴어 요리 이름  
레시피  
  
...
```

충족해야 할 조건은 아래와 같다.

- 라틴어 요리 이름에서 밑줄(\_)을 제거한다.
- 여러 개의 행으로 이루어진 레시피는 1개의 행으로 합친다.
- 요리와 요리 사이에 줄바꿈 문자를 두 개 넣어서 출력한다.