

## 언어 분석을 위한 프로그래밍(2) 중간고사 (35점)

◆ 다음을 읽고, 물음에 답 하세요. 답안은 반드시 e-class에 upload 되어 있는 답안지 양식에 작성하세요. (총 5점)

1. global variable (전역변수)와 local variable (지역변수)에 대해 간략히 서술하세요. (2점)
2. 매개변수, 기본값 매개변수, 키워드 매개변수의 정의와 특징에 대해 예시와 함께 설명하세요. (3점)

◆ 실습 문항입니다. 다음을 읽고, 각 문항의 문제를 수행할 수 있도록 파이썬 코드를 작성하세요. (총 30점)

1. 영어 단어의 글자수를 세는 사용자 정의 함수 만들고자 한다. (5점)
  - 결과값이 있는 형태와 결과값이 없는 형태로 각각 2개의 사용자 정의 함수를 만들고, 각각을 "word\_length1"과 "word\_length2"로 정의
  - input()을 통해 영어 단어를 입력 받기 (코드 작성 시 test 단어는 apple로 하세요)
  - 답안 코드 내에 각 함수 호출 부분까지 포함해서 작성하고, "단어: 글자수" 형태로 콘솔에 출력
2. 텍스트에서 단어를 추출해내는 사용자 정의 함수 "word\_extraction"를 만들고자 한다. (5점)
  - text = "Gentleness and kind persuasion win where force and bluster fail."
  - 소문자 변환, 문장기호 삭제, 중복단어 제거, 알파벳 순으로 정리
  - 답안 코드 내에 각 함수를 호출 부분까지 포함해서 작성하고, "번호: 단어" 형태로 출력 (아래 출력형태 참고)

출력 형태
1: and 2: bluster 3: fail 4: force 5: gentleness 6: kind 7: persuasion 8: where 9: win

3. 아래의 내용을 수행할 수 있는 코드를 작성하세요. (5점)  
이름, 나이, 성별 정보가 들어 있는 per\_information.txt에서 필요한 정보만 뽑아내고 출력하고자 한다. (파일 내 정보를 사전에 반드시 확인하고 충분히 숙지한 후 코드를 작성할 것!)
  - 파일을 파이썬에서 열고 읽기
  - 파일 내 정보 중, 이름만 콘솔창에 출력하고, "result\_per\_info.txt"에도 저장하기 (한 줄씩 줄바꿈하기, 아래 파일 저장 형태 참고)

파일 저장 형태
Mike John Mary Olivia Emma

- 파일을 읽고 쓸 때, **반드시** 상대경로로 써서 답안 제출!!

4. 다음을 수행할 수 있는 코드를 작성하세요. (5점)

Aesop\_001\_Task2\_009.TextGrid.utf16 파일에는 어떤 화자의 발화에 대한 시간정보와 단어, 음소 정보가 들어 있다. (파일 내 정보를 사전에 반드시 확인하고 충분히 숙지한 후 코드를 작성할 것!)

- 파이썬에서 파일을 열고 읽기
- 단어만 뽑아내서 다음과 같은 형태로 “Why do you always misunderstand?” 한 줄에 띄어쓰기 맞춰서 출력
- 파일을 읽을 때, **반드시** 상대경로로 써서 답안 제출!!

5. 드라마 “Friends”의 등장인물의 언어 사용에 대해 분석해 보고자 한다. 본 문항에 사용할 데이터는 드라마 Friends의 시즌 1 에피소드 1의 대본이다. Monica, Joey, Chandler, Phoebe, Ross, Rachel가 어떤 단어를 많이 사용하는지 분석해 보자. (10점)

- friend\_1\_1.txt 파일을 파이썬에서 열고 읽기 (파일 내 정보를 사전에 반드시 확인하고 충분히 숙지한 후 코드를 작성할 것!)
- 파일을 읽을 때, **반드시** 상대경로로 써서 답안 제출!!
- 각 등장인물이 사용한 상위 5개의 단어를 추출
  - 모든 단어 소문자로 바꾸기 (1인칭 대명사 I 포함)
  - 문장부호를 공백 “ ”으로 변환
    - 변환해야 할 문장부호: 마침표(.) 쉼표(,) 물음표(?) 느낌표(!) 하이픈(-) 작은따옴표 또는 어퍼스트로피(') 큰 따옴표(“ ”)
  - 각 등장인물의 단어 사용 빈도를 딕셔너리를 이용해 “단어: 빈도” 형태로 저장
  - Top 5 단어 뽑고 출력 (주의: 하이픈과 어퍼스트로피를 공백 “ ”으로 변환하여 두 단어로 만들었으므로 (예: don't -> don t) 정확도가 높은 분석은 아님)
    - 등장인물 별로 Top 5 단어를 추출하는 6번의 동일한 반복 작업에 대해 loop 또는 함수를 통해 처리하는 것과 그렇지 않은 경우, 점수에 차등 있음
    - Top 5 단어를 추출하기 위해서는 딕셔너리 sorting이 필요함! 아래 코드 참고!
 

```
import operator
tuple(sorted(딕셔너리변수이름.items(), key=operator.itemgetter(1), reverse=True))
```
    - 딕셔너리 sorting이 어려운 경우 딕셔너리 외 다른 자료형을 사용해서 “단어: 빈도”를 정리하고 top5를 추출해도 무관!!
  - 아래 출력형태와 같이 콘솔창에 출력

출력 형태
Top 5 words of MONICA you : 37 i : 36 s : 20 it : 16 is : 15 Top 5 words of JOEY you : 29 a : 20 the : 16 i : 15 s : 12 ... 이하 생략