

[숙제07] 한글 처리 모듈 확장

언어와 컴퓨터

2018년 10월 26일 금요일 13시까지

- 소스코드 스크립트 `hangul.py` 및 보고서 `hw07_000000.pdf` 파일을 `hw07_000000.zip` 파일로 압축하여 제출하라.
- `hangul.py` 파일에 기존에 만든 함수가 들어 있으면 앞으로 한글 처리 모듈로 활용하기 좋다. 없어도 괜찮다.
 - [숙제05] `ishangul()`, `hasbatchim()`, `get_ga()`
 - 10월 17일자 10강 실습 코드 `jamo_name()`, `conjoin()`
- 보고서는 문제 해결 방법, 코드 설명, 테스트 실행 결과 등을 포함하여 작성하라.

1 한글 처리 모듈 확장하기

파이썬 내장 함수 `ord()`와 `chr()`의 도움말 및 10월 17일자 10강 강의 내용을 참조하여 다음의 두 함수를 작성하고 문제에 답하라.

1.1 `decompose(syllable)` (3점)

매개변수 `syllable`의 문자열 인자가 한글 1음절(유니코드 한글 음절 영역의 문자, `HANGUL SYLLABLE *`)이면 초성, 중성, 종성(유니코드 한글 호환성 영역의 문자, `HANGUL LETTER *`)으로 분해한 문자열을 반환한다. 한글 1음절 이외의 경우는 예외로 간주하고 인자를 그대로 반환한다.

| | | | |
|--|--|--|--|
| <pre>>>> decompose('사') 'ㅅㅏ'</pre> | <pre>>>> decompose('ㅅ') 'ㅅ'</pre> | <pre>>>> decompose('驀') '驀'</pre> | <pre>>>> decompose('2') '2'</pre> |
| <pre>>>> decompose('숨') 'ㅅ_ㅓ'</pre> | <pre>>>> decompose('사슴') '사슴'</pre> | <pre>>>> decompose('霍') '霍'</pre> | <pre>>>> decompose('Q') 'Q'</pre> |

1.1.1 문제(1점)

위의 예시에 사용된 문자열 중에서는 10강 강의 실습 코드를 그대로 적용했을 때 예상 밖의 결과가 나오는 것이 있다. 그 원인이 무엇인가?

1.2 `compose(hangul)` (6점)

매개변수 `hangul`의 문자열 인자가 한글 호환성 영역의 초성-중성(-종성)의 연쇄로 이루어져 있으면 여기에 대응하는 **한글 음절 문자**를 반환한다. 이외의 경우는 예외로 간주하고 인자를 그대로 반환한다.

| | | | |
|---|--|--|--|
| <pre>>>> compose('ㅇㅍ') '유'</pre> | <pre>>>> compose('ㅂㅌㅇ') '병'</pre> | <pre>>>> compose('ㅇ') '이'</pre> | <pre>>>> compose('詢') '詢'</pre> |
|---|--|--|--|

주의 10강 강의 실습 코드에서 만든 `conjoin()`의 반환값은 한글 자모 영역의 문자열이다. 이 문자열은 초성 문자, 중성 문자(종성 문자)로 이루어져 있다. 반면, `compose()` 함수가 반환해야 할 것은 길이 1인 한글 음절 영역의 문자이다.

| | |
|---|---|
| <pre>>>> conjoin('ㅎㅏㅓ') '한' >>> len(conjoin('ㅎㅏㅓ')) 3</pre> | <pre>>>> compose('ㅎㅏㅓ') '한' >>> len(compose('ㅎㅏㅓ')) 1</pre> |
|---|---|