

2018학년도 2학기 언어와 컴퓨터

제11강 한글 처리 모듈

박수지

서울대학교 인문대학 언어학과

2018년 10월 17일 수요일

오늘의 목표

- 1 try/except 문으로 예외를 처리할 수 있다.
- 2 유니코드 코드 포인트 값을 사용하여 한글 음절을 구성하는 자모를 알아낼 수 있다.
- 3 유니코드 이름을 사용하여 옛한글 자모를 결합할 수 있다.

지난 시간에 배운 것

- 1 부호화와 문자 인코딩의 개념
- 2 유니코드 이전의 인코딩 방식
- 3 한글 인코딩의 문제
- 4 유니코드의 특징과 구성

```
1 def myint(string):  
2     return int(string)  
3  
4 myint(42)  
5 myint('42')  
6 myint('4.2') # ValueError  
7 myint([42]) # TypeError  
8 myint('42')) # SyntaxError
```

오류의 유형

문법 오류 문법이 잘못되어 실행이 되지 않는 오류

- `SyntaxError`, `IndentationError`

예외 문법상 올바르지만 실행 중에 감지되는 오류

- `TypeError`, `ValueError` 등 대부분의 `*Error`

예외 처리 구문

try/except

형식

```
1 try:
2     <예외가 발생할 가능성이 있는 코드>
3 except:
4     <예외가 발생했을 때 실행할 코드>
```

실습 코드: try.py

유니코드 한글 음절 영역

- 음절 하나가 문자 하나가 된다.
- 초성, 중성, 종성 순서대로 & 빠짐없이 나열된다.
- 옛한글은 포함되지 않는다.

유니코드 한글 자모 영역

- 자모 하나가 문자 하나가 된다.
- 옛한글을 포함한 초성, 중성, 종성이 각각 순서대로 나열된다.
- 초성, 중성, 종성을 연달아 쓰면 음절처럼 보인다.

유니코드 한글 음절 영역의 재구성

몇 번째 표 (초성) 몇 번째 행 (중성), 몇 번째 열 (종성)에 있는가?

초성		→ 중 성					
		ㄱ	0	1	2	...	26
↓ 중 성	0	가	각	갉	...	갈	갈
		0	1	2		26	27
	1	개	객	괘	...	괘	괘
		28	29	30		54	55
	2	가	각	갉	...	갈	괘
		56	57	58		82	83

	20	기	긱	귀	...	깊	김
		560	562	563		586	587

		→ 중 성					
초 성	ㄲ	0	1	2	...	26	27
	0	까	깍	깎	...	깟	깡
		0	1	2		26	27
	1	개	깹	깺	...	깻	깼
		28	29	30		54	55
	2	까	깍	깎	...	깟	깡
		56	57	58		82	83

	20	끼	깁	깑	...	깒	깓
		560	562	563		586	587

유니코드 한글 자모 영역에서 자모 찾기

- 코드 포인트의 규칙성을 활용하기 어렵다.
- 문자의 이름을 활용한다.

실습 코드: `hangu1.py`