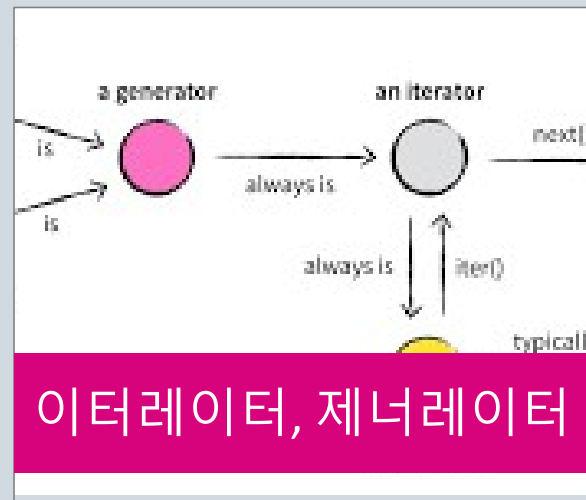
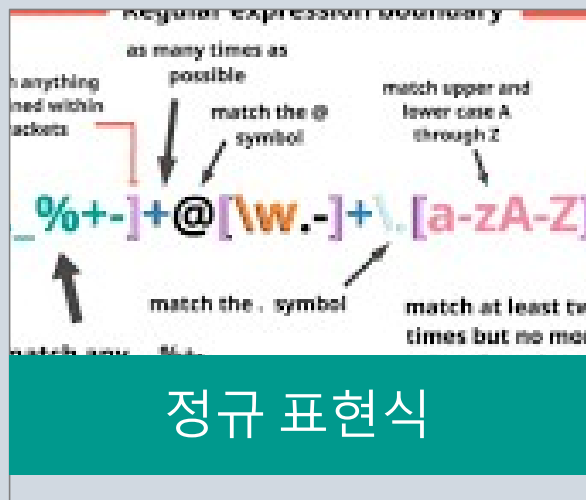


5주차 스터

목차



메타 문자

구분	메타문자	의미
한 문자	[]	문자 클래스 생성
	.	\n 외 모든 문자
	?	문자가 있거나 없거나
		좌우 문자 중 하나가 있는지
반복	{n, m}	앞의 문자가 n~m개 존재 가능
	*	앞의 문자가 0~개 존재 가능
	+	앞의 문자가 1~개 존재 가능
위치	^	문자열의 시작
	\$	문자열의 끝
그룹	()	안의 내용을 한 문자처럼 묶음

[] 메타문자

메타 문자	의미	예시
-	앞 문자부터 끝 문자까지 범위	[a-z], [A-Z], [a-zA-Z], [0-9]
^	not	[^0-9]

자주 사용하는 문자 클래스

문자 클래스	별도 표기법	의미
[0-9]	\d	숫자만 가능
[^0-9]	\D	숫자 불가능
[a-zA-Z0-9_]	\w	문자 + 숫자 + _ 가능
[^a-zA-Z0-9_]	\W	문자 + 숫자 + _ 불가능
[\t\n\r\f\v]	\s	화이트스페이스 가능
[^ \t\n\r\f\v]	\S	화이트스페이스 불가능

Re 모듈

re모듈

- 파이썬에서 정규 표현식을 지원하는 모듈

사용 방법

- 객체명 = re.compile('정규식')
- 객체명.메서드

메서드

Method	목적
<code>match(['정규식'], '문자열')</code>	문자열의 처음부터 조사하여 <code>match</code> 객체 반환
<code>search('문자열')</code>	문자열 전체에서 일치하는 문자열 조사하여 <code>match</code> 객체 반환
<code>findall('문자열')</code>	일치하는 모든 문자열을 리스트로 반환
<code>finditer('문자열')</code>	일치하는 모든 문자열을 반복 가능한 객체로 반환

match 객체 메서드

method	목적
group	매치된 문자열 반환
start	매치된 문자열의 시작 위치 반환
end	매치된 문자열의 끝 위치 반환
span	매치된 문자열의 (시작, 끝)을 튜플로 반환

컴파일 옵션

옵션	의미
DOTALL, S	메타 문자 . 이 \n를 포함 모든 문자
IGNORECASE, I	대소문자에 관계없음
MULTILINE, M	^, \$ 메타 문자가 여러 줄과 매치
VERBOSE, X	verbose 모드를 사용 정규식의 가독성을 높게 작성 가능

역슬래시 문제

문제

\ 를 문자가 아닌 메타 문자로 인식

- 정규식: `' \ asdf'`

방법1

이스케이프 처리

- 정규식: `' \ \ asdf'`

방법2

raw string 표현법 사용

- 정규식: `r' \ asdf'`

이터레이터

정의

- Next() 함수 호출 시 다음 값을 반환하는 객체

이터레이터 생성

Function

- `iter()`: 반복 가능한 객체를 이터레이터로 변환

Class

- `__iter__`: 이터레이터 객체를 반환하는 메서드
- `__next__`: `next()` 함수 호출 시 실행되는 메서드

제너레이터

정의

- 이터레이터를 생성해 주는 함수

제너레이터 생성

Function

- 차례대로 결과 반환을 위해 return 대신 yield 사용

제너레이터 표현식

- `(i * i for i in range(1, 1000))`

사용방법

next()

- 반환 값 없으면 StopIteration 예외 발생

For문

- 반환 값 없으면 for문 종료

쓰임새

이터레이터

- 복잡한 행동일 경우 사용

제너레이터

- 간단한 행동일 경우 사용
- 느긋한 계산법

문제

정규표현식과 이터레이터를 활용한 단어 필터링

문제: 텍스트 파일을 읽어 특정 패턴(예: ing로 끝나는 단어)을 찾고, 이를 이터레이터를 사용하여 한 줄씩 반환하는 프로그램을 작성하시오.

클래스 WordFilter를 작성하여 `__iter__()`를 구현하고 이터레이터를 활용
re 모듈을 이용하여 정규표현식으로 단어 필터링