

07-1 ~ 07-2 정규 표현식 시작1


Regular Expressions(정규 표현식)

복잡한 문자열을 처리할 때 사용하는 기법(파이썬만의 고유 문법X)

+ 정규표현식 참고하기 좋은 사이트)

불규칙 속에서 규칙을 찾아내는 정규 표현식

개발자들은 자연어로 주어지는 문제 상황을 파악하고 프로그램을 설계하고 작성하는 사람들이다. 이런 업무를 수행하기 위해 개발자들은 불규칙하게 쏟아지는 정보들 속에서 필요한 부분들을 걸러내거나 무

 <https://evan-moon.github.io/2020/07/24/about-regular-expression/>

`/(\w)\1{3}/`
abc **dddd** efg

- 정규 표현식을 알아야하는 이유: 코드가 상당히 간결해짐, 찾으려는 문자열이나 바꾸어야 할 문자열의 규칙이 매우 복잡할 수록 정규식의 효용은 더 커짐
- 정규 표현식의 기초, 메타 문자(meta characters)
 - meta characters란: 원래 그 문자가 가진 뜻이 아닌 **특별한 용도**로 사용하는 문자
 - example: . ^ \$ * + ? { } [] \ | ()

각 메타 문자의 의미와 사용법 알아보기

▼ 문자 클래스 []

문자클래스[]로 만들어진 정규식의 의미: '[] 사이의 문자들과 매치'

즉, 정규표현식 [abc]는 'a,b,c중 한 개의 문자와 매치'라는 의미

정규 표현식[abc]와 "a", "before", "dude"가 어떻게 매치되는지 표를 통해 살펴보자

정규식	문자열	매치 여부	설명
[abc]	a	Yes	"a"는 정규식과 일치하는 문자인 "a"가 있으므로 매치
[abc]	before	Yes	"before"는 정규식과 일치하는 문자인 "b"가 있으므로 매치
[abc]	dude	No	"dude"는 정규식과 일치하는 문자인 "a,b,c"중 어느 한 문자도 포함하고 있지 않으므로 매치X

- 문자 클래스[] 사이의 " - " 기호의 의미는 문자 사이의 범위(from to)를 의미함
ex) [a-c] = [abc], [0-5] = [012345]

- [a-zA-Z]: 알파벳 모두, [0-9]: 숫자

\$ 주의할 점: 문자 클래스 안에는 어떤 문자나 메타 문자도 사용 가능하지만, ^ 메타문자는 주의해야함.

^메타문자 의미: 반대(not), ex) [^0-9]는 숫자가 아닌 문자만 매치됨

▼ Dot (.)

Dot(.) 메타문자는 줄바꿈 문자(\n)를 제외한 모든 문자와 매치됨을 의미함

ex) a.b는 a와 b사이에 줄바꿈 문자(\n)를 제외한 어떤 문자가 와도 모두 매치됨을 의미함(a + 모든 문자 + b)

정규식	문자열	매치여부	설명
a.b	aab	Yes	가운데 존재하는 'a'는 모든 문자 라는 개념에 속하기 때문에 정규식과 매치
a.b	a0b	Yes	'0'이 모든 문자에 속하므로 정규식과 매치
a.b	abc	No	'a'와 'b'사이에 어떠한 문자라도 하나가 존재해야하므로 정규식과 매치X

- a[.]b: "a.b"문자열과 매치되고, "a0b"라는 문자열과는 매치X

▼ 반복(*)

반복(*)은 * 바로 앞에 있는 문자가 0부터 무한대로 반복될 수 있다는 의미

ca*t: "a"가 0번 이상 반복되면 매치됨

정규식	문자열	매치여부	설명
ca*t	ct	Yes	"a"가 0번 반복되어 매치
ca*t	cat	Yes	"a"가 1번 반복되어 매치(0번 이상)
ca*t	caaat	Yes	"a"가 3번 반복되어 매치(0번 이상)

▼ 반복(+)

반복을 나타내는 또다른 메타문자(+)

+는 최소 1번 이상 반복될 때 사용(*는 0번 이상 반복)

ca+t: 메타문자+ 앞에 있는 "a"가 최소 1번 이상 반복되어야 매치

정규식	문자열	매치여부	설명
ca+t	ct	No	“a”가 0번 반복되어 매치되지 않음(1번 이상)
ca+t	cat	Yes	“a”가 1번 반복되어 매치
ca+t	caaat	Yes	“a”가 3번 반복되어 매치

▼ 반복({m,n},?)

단순히 몇 회 이상 반복이 아니라 m회부터 n회까지 반복으로 제한하고 싶을 경우 사용

1. {m}: 반드시 m회 반복한 문자열과 매치

ca{2}t: “a”가 2번 반복되면 매치

2. {m,n}: m회에서 n회 반복된 문자열과 매치

ca{2,5}t: “a”가 2회에서 5회 반복된 문자열과 매치

3. ? : 메타문자?가 의미하는 것은 {0,1}

ab?c: “b”가 0회에서 1회 반복되면 매치(“b”가 있어도 되고 없어도 된다)

정규식	문자열	매치여부	설명
ab?c	abc	Yes	“b”가 1번 사용되어 매치
ab?c	ac	Yes	“b”가 0번 사용되어 매치