

정규 표현식

[Regular Expression]

Kim Hye Kyung
topickim@naver.com

정규 표현식[Regular Expression]이란?

특정 패턴의 문자열을 표현 및 가공하기 위한 기법

정규 표현식[Regular Expression]이란?

- 정교한 문자열 매칭 기능 제공
- 정규 표현식 또는 정규화라고도 함
- 데이터 가공시 간편하고 직관적인 코드 작성이 가능
 - 이메일 주소, URL, 전화 번호 처럼 보이는 문자열 표현 및 검색에도 사용
- 문자열 교체에 필요한 기능 보유
 - 문자열 속에서 부분 문자열을 찾는 일이며, 찾은 부분을 교체하기도 함
 - 예시
 - 이메일 주소처럼 보이는 문자열을 찾아 그 주소를 하이퍼링크형으로 변경 가능

정규 표현식 문자열 판단하기

- 필요 모듈

- re 모듈

- regular expression의 약어

- 문자열 판단하기

- match 함수

- 정규표현식 패턴과 판단할 문자열 적용
 - 문자열 처음부터 매칭되는지 판단

- search 함수

- 문자열이 맨 앞에 오는지? 맨 뒤에 오는지 판단
 - ^문자열 - 문자열이 맨 앞에 오는지 판단
 - 문자열\$ - 문자열이 맨 뒤에 오는지 판단

```
re.match('패턴', '문자열')
```

```
re.search('패턴', '문자열')
```

정규 표현식 문자열 판단하기

- 정규 표현식의 주요 메서드
 - 문자열 객체의 메서드와 함께 사용하는 것이 일반적
 - 정규 표현식을 사용하는 String 객체의 메서드

메서드명	설 명
match(regExp)	문자열 처음부터 매칭되는지 판단
search(regExp)	문자열 일부분이 매칭되는지 판단

정규 표현식 문자열 판단하기[예시]

- 지정된 문자열이 하나라도 포함되는지 판단
 - 기본 개념
 - OR 연산자와 같음
 - 문법
 - 'private|public' 문자열에 private 또는 public이 포함되었는지 판단
 - 'private|defalut|public' 문자열에 private 또는 default 또는 public 이 포함되었는지 판단

정규 표현식 주요 기호

참조 사이트 : https://www.w3schools.com/jsref/jsref_obj_regexp.asp

의미	내용
.	임의의 한 문자가 존재
?	문자가 존재하거나 존재하지 않음(0, 1)
*	문자가 미존재 또는 무한대로 존재(0~*)
+	문자가 한번 이상 존재(1~*)
^	기호 바로 뒤의 문자로 문자열이 시작
\$	기호 바로 앞의 문자로 문자열이 끝남
{숫자}	숫자만큼 반복
{숫자,}	숫자 이상만큼 반복
{숫자1, 숫자2}	숫자 1 이상, 숫자 2 이하 만큼 반복
(문자열)	문자나 문자열을 묶음
[문자1, 문자2,...]	대괄호 안에 있는 문자들이 존재하는지 검색
[^]	^ 기호 뒤 문자가 존재하지 않음

정규 표현식 주요 기호

의미	내용
[:알파벳:]	알파벳만 검색
[:알파벳숫자]	알파벳과 숫자만 검색
[:숫자:]	숫자만 검색
[:upper:]	대문자만 검색
\w	\w 역슬래시 글자 자체를 검색
\d	모든 숫자를 검색, [0-9]와 동일
\D	[^0-9]와 동일, 숫자를 제외한 모든 문자를 검색
\s	공백을 검색
\S	공백이 아닌 문자를 검색

정규 표현식 주요 기호

의미	내용
\w	숫자 또는 문자를 검색 [a-zA-Z0-9] word 를 표현하며 알파벳 + 숫자 + _ 중의 한 문자임을 의미
\W	숫자 또는 문자가 아닌 것을 검색 [^a-zA-Z0-9] non word를 표현하며 알파벳 + 숫자 + _ 가 아닌 문자를 의미
[xy]	문자 선택을 표현하며 x 와 y 중에 하나를 의미 x와 y 중에 먼저 나온 단어 검색
[^xy]	not 을 표현하며 x 및 y 를 제외한 문자를 의미
[x-z]	range를 표현하며 x ~ z 사이의 문자를 의미
\w^	escape 를 표현하며 ^ 를 문자로 사용함을 의미
\wb	word boundary를 표현하며 문자와 공백 사이의 문자를 의미
\WB	non word boundary를 표현하며 문자와 공백 사이가 아닌 문자를 의미
\wd	digit 를 표현하며 숫자를 의미
\WD	non digit 를 표현하며 숫자가 아닌 것을 의미

그룹 사용하기

- 정규 표현식 그룹은 해당 그룹과 일치하는 문자열을 얻어 올 때 사용
- 예시
 - 공백으로 구분된 숫자를 두 그룹으로 나눠서 찾은 뒤 각 그룹에 해당하는 문자열(숫자)를 가져옴

```
1 #p = re.compile('(test OK)') #p = re.compile('(test)+')
2 m = re.match('([0-9]+) ([0-9]+)', '150 22')
3 print(m)
4 type(m)
```

<re.Match object; span=(0, 6), match='150 22'>

re.Match

```
1 print(m)
```

<re.Match object; span=(0, 6), match='150 22'>

```
1 # 해당 문자열을 한번에 반환
2
3 print(m.group(0))
```

150 22

```
1 # 첫 번째 그룹에 매핑된 문자열 반환
2
3 print(m.group(1))
```

150

```
1 # 두 번째 그룹에 매핑된 문자열 반환
2
3 print(m.group(2))
```

22