

@@@ 정규 표현식 @@@

#정규 표현식이란?

정규식, regex, regexp라고도 부른다. 특정한 규칙을 가진 문자열의 집합을 표현하는데 사용하는 형식언어. 주로 프로그래밍 언어나 텍스트 에디터 등에서 문자열의 검색과 치환을 위해 사용. 정규표현식에서 사용하는 기호는 Meta문자라고 한다. 표준은 POSIX이며 확장 방식 중 대중적인 것은 Perl과 PCRE.

#정규식 메타 문자

메타 문자	설명	예시
.	아무 문자	
^	문자열의 시작. [] 안에서는 부정의 의미.	^http
\$	문자열의 끝	png& 하면 png를 다 찾겠지.
\b	단어의 경계. 공백, 탭, 콤마, 대시 등	\bplay\b 하면 play라는 단어를 검색하겠지
\B	\b가 아닌 것.	\Bplay\B 하면 play는 안하고 playground, playball 매치함.
\s	공백문자	
\S	공백문자가 아닌 것	
\d	숫자. [0-9]랑 동일	
\D	숫자가 아닌 것. [^0-9]랑 동일	
\w	Word를 만들 수 있는 것들. 알파벳, 숫자	
\W	\w가 아닌 것	
\n	개행	
\r	캐리지리턴	
\t	Xo	
\w	문자 자체로 쓸 때.	\w, \w* \w\$ \w(등등

#선택 패턴

A B	A혹은 B 둘다 매칭	pdf png
[abc]	a혹은 b혹은 c	[cbh]at 하면 cat bat hat
[^abc]	abc빼고	[^cbh]at하면 rat eat 등등
[a-z]	a~z까지	한글은 [ㄱ-힣]

#그룹

괄호로 둘러싸면 그룹이 되어 캡처한다. 하나로 묶이는 놈들 처리할 때 사용
ex)

(tom|pot)ato -> tomato, potato

(a|i){3}bc -> aaabc, aaibc, aiabc 등등

(to)maW1 -> W1로 (to)를 호출할 수 있음 -> tomato

(a|b|c){2}maW1 -> aaaaaa, bbmabb 등등

#비캡처링 그룹

(?:)하면 그룹핑을 하긴 하는데 캡처는 안한다. 캡처하고 사용 안하면 되긴 함.

#수량 한정자

동일한 글자 혹은 동일한 족(family)이 n개 만큼 나오는 경우에 수량한정자를 뒤에 붙일 수 있다.

?	앞의 표현식이 없을 수도 있다.	nonoks?에서 s?는 있을 수도 있고 없을 수도 있다. nononk일 수도 있고 nonoks일 수도 있다. no(nok)?이면 no 일 수도 있고 nonok일 수도 있다.
*	0개 이상이다.	nWd* -> n 뒤에 Wd(숫자)가 0개 이상. -> n, n12, n135214 등등
+	1개 이상이다.	nWd+ -> n(x) n1, n2, n56, n1356
{n}	n개 있다.	nWd{n} -> n123, n654, n432. 딱 n개만
{n, m}	n개 이상 m개 이하	
{n,}	n개 이상	{n}은 n개 이하

*, +는 greedy하게 동작한다. 가능한 많은 글자를 다 먹음. ?와 조합하면 반대로 최소한만 먹는다.
ex)

I like apples and bananas

^.*s -> ^처음부터 .아무 문자나 *n개 이상 있고 s로 끝난다. -> 전체 문장

^.*?s -> ^처음부터 .아무문자나 *?s s로 끝나는데 까지 자른다. -> I like apples

#예제

1. 개행이 아닌 문자 배치

abc.def.ghijklx의 형태에 매치하는 패턴. 각 변수 a-x는 알파벳일 필요 없고 공백 제외 아무 문자를 의미한다.

-> 비공백문자 3개 `[^Wn]{3}`

-> 비캡처 그룹 시작 `(?: -> . 들어감 (?W. -> 다시 문자 3개 (?W.[^Wn]{3}) - 애를 3번`

답: `[^Wn]{3}(?:W.[^Wn]{3}){3}`

2-1. 휴대폰 번호 매치

010 8435 1989, 01084351989, 010.8435.1989, 010-8435-1989, 84351989 등등

-> `(010)?` 010으로 시작할 수도 있고 안할 수도 있음

-> `WD?` 숫자는 아닌 무언가가 들어갈 수도 있고 없을 수도 있음

-> `Wd{4}` 숫자 4개

-> `(WD?Wd{4}){2}` 그 패턴이 2번

답: `(010)?(WD?Wd{4}){2}`

2-2. 휴대폰 번호 매치

옛날폰은 010도 쓰고 016, 017등 쓴다. 그리고 중간이 3자리다.

-> `01[016789]` 010일수도 있고, 016일수도 등등

-> `WD?` 숫자는 아닌 거(.이나 -) 하나가 들어갈 수도 있고 안 들어갈 수도 있음.

-> `Wd{3,4}` 요즘 건 4자리인데 3자리일 수도 있다.

-> `WD?` 또 한번 더

-> `Wd{4}` 끝자리

답: `01[016789]WD?Wd{3,4}WD?Wd{4}`

<https://regexone.com/lesson/> 그 외 문제들

<https://regexr.com/> <https://regex101.com/> 테스트 사이트

#추가 내용

POSIX – 문자클래스			
<code>[:alnum:]</code>	알파벳과 숫자	<code>[:alpha:]</code>	알파벳
<code>[:digit:]</code>	0~9 사이	<code>[:lower:]</code>	알파벳 소문자
<code>[:blank:]</code>	탭과 공백문자	<code>[:upper:]</code>	알파벳 대문자

@@@ 쿼리 스트링 @@@

사용자가 입력한 데이터를 서버를 전달하는 방법.

GET 방식으로 요청했을 때 URL 주소 뒤에 입력 데이터를 함께 제공하는 방법으로
"리소스?이름=값"의 형식을 취한다.

쿼리 스트링은 브라우저의 URL 의 주소 입력란에서 확인할 수 있다. ?로 시작한다. 쿼리 스트링이
두개 이상일 경우에는 &로 쿼리 스트링을 연결한다.

:포트/앱/?id=nonok&pw=nonok0812 이런 식으로 ㅇㅇ