

3. 정규표현식을 사용한 입력값 검사

(1) 정규 표현식 (Regular expression)

- 특정한 규칙을 가진 문자열의 집합을 표현하는데 사용하는 형식언어이다.
- 많은 텍스트 편집기와 프로그래밍 언어에서 문자열의 검색과 치환을 위해서 사용되고 있다.
- 자바에서도 정규표현식을 J2SE 1.4부터 지원하기 시작했으며, 관련된 주요 클래스들은 `java.util.regex` 패키지에 포함되어 있다.

* 정규표현식 표현방법

정규표현식은 표준인 POSIX의 정규표현식과 POSIX 정규표현식에서 확장된 Perl방식의 PCRE가 대표적이며, 이외에도 수많은 정규표현식이 존재하며 정규표현식 간에는 약간의 차이점이 있으나 거의 비슷하다. 정규표현식에서 사용하는 기호를 Meta문자라고 한다. Meta문자는 표현식 내부에서 특정한 의미를 갖는 문자를 말하며, 공통적인 기본 Meta문자의 종류로는 다음과 같다.

- 일반적으로 정규표현식은 /와 / 사이에 위치한다.
- 경우에 따라서는 /와 /는 생략해서 사용되기도 한다.

식	기능	설명
.	문자	한 개의 어떤 문자와도 일치
\d	숫자	한 개의 숫자와 일치
\w	문자와 숫자	한 개의 문자나 숫자와 일치
\s	공백문자	공백, 탭, 줄 바꿈, 캐리지 리턴 문자와 일치
^	시작	패턴의 시작을 표시
\$	끝	패턴의 끝을 표시
[]	문자종류, 문자범위	[abc]는 a 또는 b 또는 c를 나타낸다. [a-z]는 a부터 z까지 중의 하나, [1-9]는 1부터 9까지 중의 하나를 나타낸다.

- 정규표현식에는 메타 문자 뒤에 수량 한정자(qualntifier)를 붙일 수 있다.
수량한정자는 문자가 몇 번 반복되느냐를 나타낸다.

수량 한정자	기능	설명
()	문자를 그룹핑한다.	"abc adc"와 "a(b d)c"는 같은 의미를 가진다.
*	0회 이상 반복	"a*b"는 "b", "ab", "aab", "aaab"를 나타낸다.
+	1회 이상	"a+b"는 "ab", "aab", "aaab"를 나타낸다.
?	0 또는 1회	"a?b"는 "b", "ab"를 나타낸다.
{m}	m회	"a{3}b"는 "aaab"를 나타낸다.

- 정규표현식을 실제로 사용할 때 언어마다 사용방법이 각각 다르다.
- <http://regexlib.com> : 많이 사용되는 정규식 패턴을 모아놓은 사이트

* 정규 표현식의 사용

- `boolean java.util.regex.Pattern.matches(String arg0, CharSequence arg1)` 메소드 사용
- `arg1`이 `arg0`의 형식에 맞는지를 검사하고 결과를 `boolean`으로 리턴한다.

* 정규 표현식의 종류

- 모든 문자
"^[]*\$"
- 숫자 모양에 대한 형식 검사
"^[0-9]*\$"
- 실수 모양에 대한 형식 검사
"^[0-9]+([.][0-9]+)?\$"
- 영문으로만 구성되었는지에 대한 형식 검사
"^[a-zA-Z]*\$"
- 한글로만 구성되었는지에 대한 형식 검사
"^[ㄱ-ㅎ가-힣]*\$"
- 영문과 숫자로만 구성되었는지에 대한 형식 검사
"^[a-zA-Z0-9]*\$"
- 한글과 숫자로만 구성되었는지에 대한 형식 검사
"^[ㄱ-ㅎ가-힣0-9]*\$"
- 이메일 형식인지에 대한 검사. "아이디@도메인"의 형식을 충족해야 한다.
"^[0-9a-zA-Z]+([_a-z0-9-]+)*@(?:\\w+\\.)+\\w+\$"
- "-"없이 핸드폰번호인지에 대한 형식검사.
"^01(?:011|[6-9])(?:\\d{3}\\d{4})\\d{4}\$"
- "-"없이 전화번호인지에 대한 형식검사. 각 부분에 대한 자리수도 충족시켜야 한다.
"^\\d{2,3}\\d{3,4}\\d{4}\$"
- "-"없이 주민번호에 대한 글자수 및 뒷자리 첫글자가 1~4의 범위에 있는지에 대한 검사
"^\\d{6}[1-4]\\d{6}\$"