

자바 정규표현식

자바의 정규표현식은 J2SE 1.4 부터 지원되지 시작했습니다. 관련된 주요 클래스들은 `java.util.regex` 패키지에 있습니다.

Pattern 클래스

`Pattern` 객체는 Perl 문법과 비슷한 형태로 정의된 정규표현식을 나타냅니다. 문자열로 정의한 정규표현식은 사용되기 전에 반드시 `Pattern` 클래스의 인스턴스로 컴파일되어야 합니다. 컴파일된 패턴은 `Matcher` 객체를 만드는 데 사용되며, `Matcher` 객체는 임의의 입력 문자열이 패턴에 부합되는지 여부를 판가름하는 기능을 담당하합니다. 또한 `Pattern` 객체들은 비상태유지 객체들이기 때문에 여러 개의 `Matcher` 객체들이 공유할 수 있습니다.

Matcher 클래스

`Matcher` 객체는 특정한 문자열이 주어진 패턴과 일치하는가를 알아보는데 이용됩니다. `Matcher` 클래스의 입력값으로는 `CharSequence` 라는 새로운 인터페이스가 사용되는데 이를 통해 다양한 형태의 입력 데이터로부터 문자 단위의 매칭 기능을 지원 받을 수 있습니다. 기본적으로 제공되는 `CharSequence` 객체들은 `CharBuffer`, `String`, `StringBuffer` 클래스가 있습니다.

`Matcher` 객체는 `Pattern` 객체의 `matcher` 메소드를 통해 얻어진다. `Matcher` 객체가 일단 만들어지면 주로 세 가지 목적으로 사용된다.

- 주어진 문자열 전체가 특정 패턴과 일치하는 가를 판단(matches).
- 주어진 문자열이 특정 패턴으로 시작하는가를 판단(lookingAt).
- 주어진 문자열에서 특정 패턴을 찾아낸다(find).

이들 메소드는 성공 시 `true` 를 실패 시 `false` 를 반환합니다.

또한 특정 문자열을 찾아 새로운 문자열로 교체하는 기능도 제공됩니다. `appendReplacement` 메소드는 일치하는 패턴이 나타날 때까지의 모든 문자들을 버퍼로 옮기고 찾아진 문자열 대신 교체 문자열로 채워 넣습니다. 또한 `appendTail` 메소드는 캐릭터 시퀀스의 현재 위치 이후의 문자들을 버퍼에 복사해 넣습니다. 다음 절에 나오는 예제 코드를 참고하도록 합시다.

CharSequence 인터페이스

`CharSequence` 인터페이스는 다양한 형태의 캐릭터 시퀀스에 대해 일관적인 접근 방법을 제공하기 위해 새로 생겨났으며, `java.lang` 패키지에 존재합니다. 기본적으로 `String`, `StringBuffer`, `CharBuffer` 클래스가 이를 구현하고 있으므로 적절한 것을 골라 사용하면 되며, 인터페이스가 간단하므로 필요하면 직접 이를 구현해 새로 하나 만들어도 됩니다.

표현식	설명
^	문자열의 시작
\$	문자열의 종료
.	임의의 한 문자 (문자의 종류 가리지 않음) 단, W 는 넣을 수 없음
*	앞 문자가 없을 수도 무한정 많을 수도 있음
+	앞 문자가 하나 이상
?	앞 문자가 없거나 하나있음
[]	문자의 집합이나 범위를 나타내며 두 문자 사이는 - 기호로 범위를 나타낸다. []내에서 ^가 선행하여 존재하면 not 을 나타낸다.
{}	횟수 또는 범위를 나타낸다.
()	소괄호 안의 문자를 하나의 문자로 인식
	패턴 안에서 or 연산을 수행할 때 사용
Ws	공백 문자
WS	공백 문자가 아닌 나머지 문자
Ww	알파벳이나 숫자
WW	알파벳이나 숫자를 제외한 문자
Wd	숫자 [0-9]와 동일
WD	숫자를 제외한 모든 문자
W	정규표현식 역슬래시(W)는 확장 문자 역슬래시 다음에 일반 문자가 오면 특수문자로 취급하고 역슬래시 다음에 특수문자가 오면 그 문자 자체를 의미
(?)	앞 부분에 (?) 라는 옵션을 넣어주면 대소문자를 구분하지 않음