

Taking a First Look at Regex Support



Victor Grazi

ORACLE JAVA CHAMPION, SPEAKER AND GEEK

@vgrazi



Introduction



The Regex Tool

Regex Basics

- Literals
- Metacharacters – The Language of Regex
- Dot – The Regex wildcard

Java Regex Support

- Native String class operations
 - matches
 - replace/replaceAll
 - split
- Pattern and Matcher classes
- PatternSyntaxException



Demo



Validate values from a web form, such as

- Address
- Postal code
- Email
- Phone

Validation algorithms for each field type

Using straight Java, requires many lines of code to parse each section





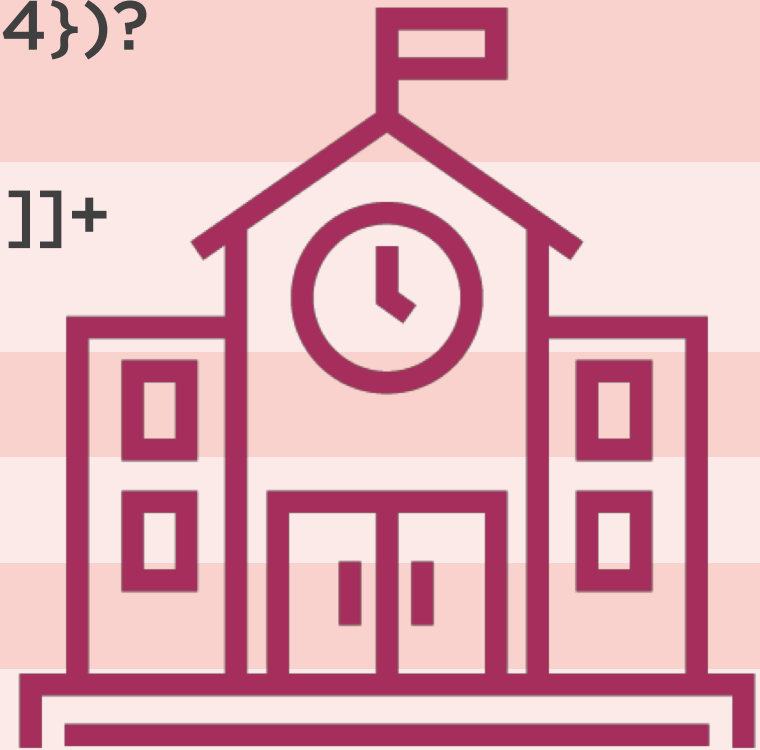
`dir my-file*.xls`



```
public void findFile() {  
    File dir = new File(".");  
    FileFilter fileFilter = new RegexFileFilter  
        ("^.*[mM]y-file(-\\d+)?\\.java$");  
    File[] files = dir.listFiles(fileFilter);  
    for (int i = 0; i < files.length; i++) {  
        System.out.println(files[i]);  
    }  
}
```



Type	Example	RegEx
Phone:	212-345-6789 or (212)-345-6789 ✓	<code>\(?:\d{3}\)? [-\s]?\d{3}-?\d{4}</code>
Date:	2019-07-2 or 2019/7/2 But not 2019/7-2 ✗	<code>\d{4}([-/])?\d{1,2}\1\d{1,2}</code>
Zip code:	11223 or 11223-4567 ✓	<code>\d{5}(-\d{4})?</code>
Non-numeric	Victor_Grazi ✓	<code>[\w&&[^\d]]+</code>
Alphabetic	AaBbCc ✓	<code>[A-Za-z]+</code>



Summary



File searching

Validation

Text searching

Text extraction

Expression reformulating

Text cleansing



String class Regex support

matches()

split()

replaceAll()

replaceFirst()

*** But not replace()**



replaceAll()

Replaces all occurrences of the supplied regex

```
"one + one = 2".replaceAll("one", "1")  
> 1 + 1 = 2
```

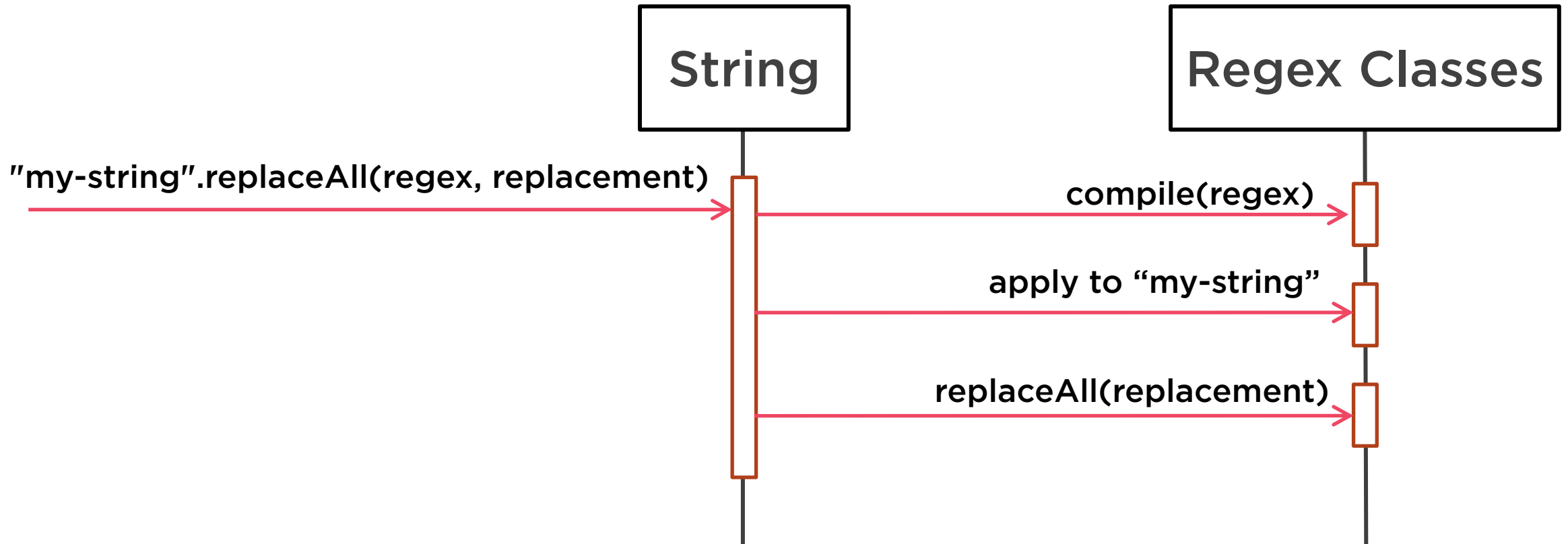
replaceFirst()

Replaces just the first occurrence of the supplied regex

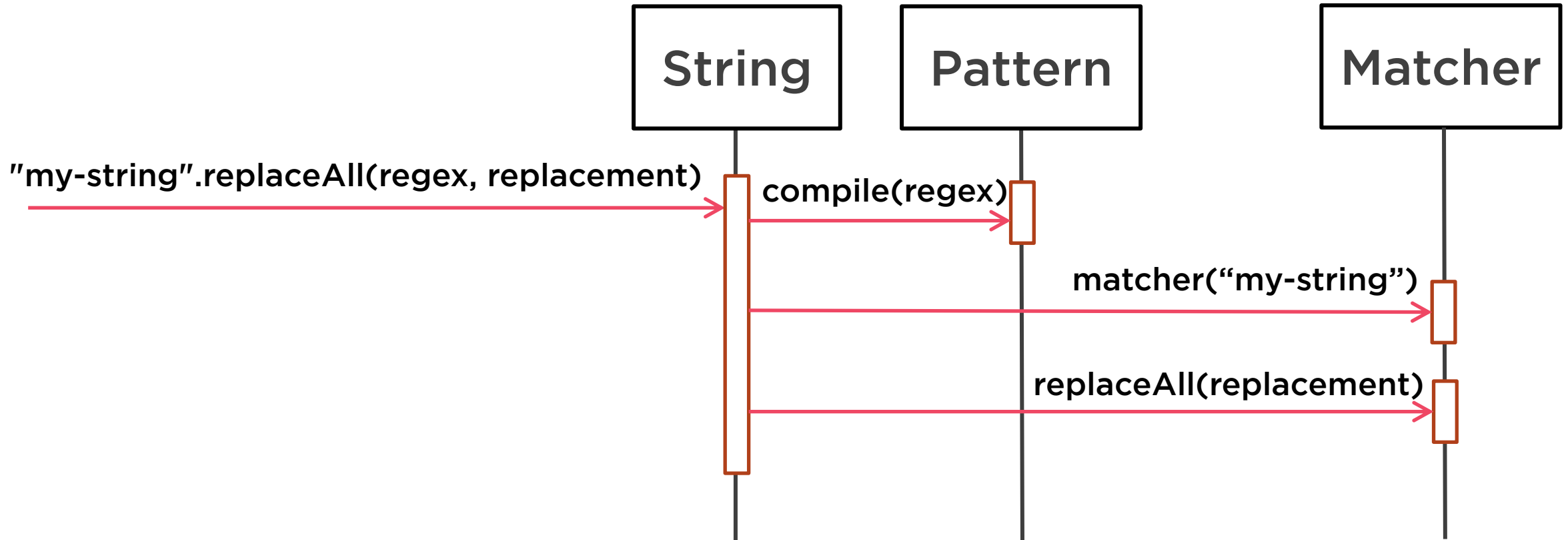
```
"one + one = 2".replaceFirst("one", "1")  
> 1 + one = 2
```



Regex Pattern Compilation



Regex Pattern Compilation



Regex Pattern Compilation

```
Pattern pattern = Pattern.compile("\\d{2}/\\d{2}/\\d{4}");
```



Validate Master Card Number

MasterCard number:

- Always 16 digits
- Either starts with 51 through 55
- Or with 2221 through 2720



Or in Regex:

```
^(?:5[1-5][0-9]{2}|222[1-9]|22[3-9][0-9]|2[3-6][0-9]{2}|27[01][0-9]|2720)[0-9]{12}$
```



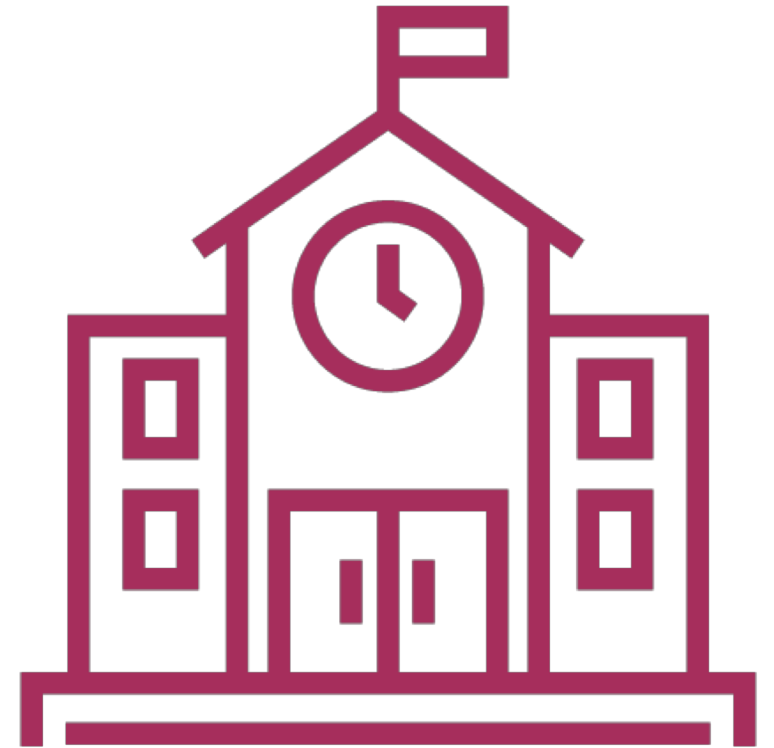
Use Pattern as a Stream Filter

```
List<String> valid = collection.stream()  
    .filter(pattern.asPredicate())  
    .collect(Collectors.toList());
```



Introducing Matcher replaceAll and replaceFirst

```
Pattern pattern = Pattern.compile("/");  
Matcher matcher = Pattern.matcher("12/21/18");  
String replaced = matcher.replaceAll("-");  
--> 12-21-18  
  
replaced = matcher.replaceFirst("-");  
--> 12-21/18  
  
matcher = Pattern.matcher(replaced);  
replaced = matcher.replaceFirst("-");  
--> 12-21-18
```



Summary



The Regex Tool

Regex Basics

- Literals
- Metacharacters – The Language of Regex
- Dot – The Regex wildcard

Java Regex Support

- Native String class operations
 - matches
 - replaceAll/replaceFirst
 - split
- Pattern and Matcher classes
- PatternSyntaxException

