Lesson 11: Match groups

Regular expressions allow us to not just match text but also to **extract information for further processing**. This is done by defining **groups of characters** and capturing them using the special parentheses **(** and **)** metacharacters. Any subpattern inside a pair of parentheses will be **captured** as a group. In practice, this can be used to extract information like phone numbers or emails from all sorts of data.

Imagine for example that you had a command line tool to list all the image files you have in the cloud. You could then use a pattern such as **^(IMG\d+\.png)$** to capture and extract the full filename, but if you only wanted to capture the filename without the extension, you could use the pattern **^(IMG\d+)\.png$** which only captures the part before the period.

Go ahead and try to use this to write a regular expression that matches only the filenames (not including extension) of the PDF files below.

Exercise 11: Matching Groups

Task	Text	Capture Groups
Capture	file_record_transcript.pdf	file_record_transcript
Capture	file_07241999.pdf	file_07241999
Skip	testfile_fake.pdf.tmp	

```
^(file_[\w\W]+[^\.pdf])
```

Continue ›

Solve the above task to continue on to the next problem, or read the Solution.

Next – Lesson 12: Nested groups (/lesson/nested_groups)
Previous – Lesson 10: Starting and ending (/lesson/line_beginning_end)

Email (mailto:regexone@gmail.com) | Twitter (http://twitter.com/regexone)

Additional Courses
Interactive SQL Lessons (https://sqlbolt.com)