Scala regexps: how to return matches as array or list

Good day.

Is there a simple way to return regexp matches as an array?
Here is how I am trying in 2.7.7:

```
val s = """6 1 2"""
val re = """(\d+)\s(\d+)\s(\d+)""".r
for (m <- re.findAllIn (s)) println (m) // prints "6 1 2"
re.findAllIn (s).toList.length // 3? No! It returns 1!
```

But I then tried:

```
s match {
  case re (m1, m2, m3) => println (m1)
}
```

And this works fine! m1 is 6, m2 is 1, etc.

Then I found something that finally completed my confusion:

```
val mit = re.findAllIn (s)
println (mit.toString)
println (mit.length)
println (mit.toString)
```

That prints:

```
non-empty iterator
1
empty iterator
```

So the "length" call somehow modifies the state of the iterator. What is going on here?

regex scala

edited May 25 '11 at 10:31 user unknown 18.8k 6 38 88

asked Jan 14 '10 at 17:35 Dfr 1,553 3 25 42

Your call to findAllIn(s) is matching the whole string, so your resulting list is not List(6 1 2), but is really List("6 1 2") of length 1 – Mitch Blevins Jan 14 '10 at 18:10

3 Answers

Ok, first of all, understand that `findAllIn` returns an `Iterator`. An `Iterator` is a consume-once mutable object. ANYTHING you do to it will change it. Read up on iterators if you are not familiar with them. If you want it to be reusable, then convert the result of findAllIn into a `List`, and only use that list.

Now, it seems you want all matching *groups*, not all matches. The method `findAllIn` will return all matches of the *full* regex that can be found on the string. For example:

```
scala> val s = """6 1 2, 4 1 3"""
s: java.lang.String = 6 1 2, 4 1 3

scala> val re = """(\d+)\s(\d+)\s(\d+)""".r
re: scala.util.matching.Regex = (\d+)\s(\d+)\s(\d+)

scala> for(m <- re.findAllIn(s)) println(m)
6 1 2
4 1 3
```

See that there are two matches, and neither of them include the ", " at the middle of the string, since that's not part of any match.

If you want the groups, you can get them like this:

```
scala> val s = """6 1 2"""
s: java.lang.String = 6 1 2

scala> re.findFirstMatchIn(s)
res4: Option[scala.util.matching.Regex.Match] = Some(6 1 2)

scala> res4.get.subgroups
res5: List[String] = List(6, 1, 2)
```

Or, using `findAllIn` , like this:

```
scala> val s = """6 1 2"""
s: java.lang.String = 6 1 2

scala> for(m <- re.findAllIn(s).matchData; e <- m.subgroups) println(e)
6
1
2
```

The `matchData` method will make an `Iterator` that returns `Match` instead of `String` .

answered Jan 14 '10 at 18:46
Daniel C. Sobral
198k 49 365 561

Microsoft
Microsoft Azure
Got 5 minutes? Then you have time to deploy your first cloud solution.
Try Azure—free

There is a difference between how unapplySeq interprets mulitple groups and how findAllIn does. findAllIn scans your pattern over the string and returns each string that matches (advancing by the match if it succeeds, or one character if it fails).

So, for example:

```
scala> val s = "gecko 6 1 2 3 4 5"
scala> re.findAllIn(s).toList
res3: List[String] = List(6 1 2, 3 4 5)
```

On the other hand, unapplySeq assumes a *perfect* match to the sequence.

```
scala> re.unapplySeq(s)
res4: Option[List[String]] = None
```

So, if you want to parse apart groups that you have specified in an exact regex string, use unapplySeq. If you want to find those subsets of the string that look like your regex pattern, use findAllIn. If you want to do both, chain them yourself:

```
scala> re.findAllIn(s).flatMap(text => re.unapplySeq(text).elements )
res5: List[List[String]] = List(List(6, 1, 2), List(3, 4, 5))
```

answered Jan 14 '10 at 18:28
Rex Kerr
119k 15 213 324

Try this:

```
  val s = """6 1 2"""
```

```
val re = """\d+""".r
println(re.findAllIn(s).toList) // List(6, 1, 2)
println(re.findAllIn(s).toList.length) // 3
```

And, if you really need a list of the match groups within a singe Regex:

```
val s = """6 1 2"""
val Re = """(\d+)\s(\d+)\s(\d+)""".r
s match {  // this is just sugar for calling Re.unapplySeq(s)
    case Re(mg@_*) => println(mg) // List(6, 1, 2)
}
```

edited Jan 14 '10 at 18:33

answered Jan 14 '10 at 18:15
Mitch Blevins
9,184 2 29 27