Parsing with PowerShell and Regular Expressions

Jeff Hicks
AUTHOR/TEACHER

@jeffhicks | https://jdhitsolutions.com

Regular Expression Parsing

Optional Matches

Named Captures

Look Arounds

```
\w+\.ps([md])?1
```

Optional Matches

Match on something that *might* exist in the string

(<your pattern>)?

```
PS C:\> $f = "a.ps1","b.bat","c.psd1","d.psm1"
```

Optional Matches

```
PS C:\> $f = "a.ps1","b.bat","c.psd1","d.psm1"
PS C:\> [regex]$rx = "\w+\.ps([md])?1"
```

Optional Matches

```
PS C:\> $f = "a.ps1","b.bat","c.psd1","d.psm1"
PS C:\> [regex]$rx = "\w+\.ps([md])?1"
PS C:\> $rx.matches($f).value
a.ps1
c.psd1
d.psm1
```

Optional Matches

You may have other regular expression choices

Named Captures

Create names for matching patterns

(?<capture-name>target-pattern)

```
PS C:\> [regex]$rx = "\\\\(?<server>)\S+\\(?<share>)\S+"
```

Named Captures

```
PS C:\> [regex\$rx = "\\\\(?<server>)\S+\\(?<share>)\S+"
PS C:\> $rx.Match("\\srv\pub-data")
```

Named Captures

```
Groups   : {0, server, share}
Success  : True
Name     : 0
Captures : {0}
Index    : 0
Length   : 14
Value    : \\srv\pub-data
```

Named Captures

Reference captures by name

Look Ahead

Match an expression by looking for something after...

...but don't include the look ahead result

target-pattern(?=regex pattern)

```
PS C:\> [regex]$rx = "\w+(?=\s\(\d+\))"
```

Look Ahead

Match on a word that is followed by text like “(40)”

I’m escaping the ( ) characters

```
PS C:\> [regex]$rx = "\w+(?=\s\(\d+\))“
PS C:\> $data = "alpha (44)","regular line","beta (123)"
```

Look Ahead

```
PS C:\> [regex]$rx = "\w+(?=\s\(\d+\))“
PS C:\> $data = "alpha (44)","regular line","beta (123)“
PS C:\> $rx.Matches($data).value
```

Look Ahead

```
PS C:\> [regex]$rx = "\w+(?=\s\(\d+\))“
PS C:\> $data = "alpha (44)","regular line","beta (123)“
PS C:\> $rx.Matches($data).value
alpha
beta
```

Look Ahead

Look Behind

Match an expression by looking for something before...

...but don't include the look behind result

(?<=regex pattern)target-pattern

```
PS C:\> [regex]$rx = "(?<=\(\d+\)\s+)\w+"
```

Look Behind

Look for “(40) ” behind the target pattern

```
PS C:\> [regex]$rx = "(?<=\(\d+\)\s+)\w+"
PS C:\> $data = "(44) alpha","regular line","(123) beta"
```

Look Behind

```
PS C:\> [regex]$rx = "(?<=\(\d+\)\s+)\w+"
PS C:\> $data = "(44) alpha","regular line","(123) beta"
PS C:\> $rx.Matches($data).value
alpha
beta
```

Look Behind

Demo

Some practical examples...

Summary

You have to know your data

Your data must be predictable and consistent

Build and test regular expression patterns outside of PowerShell

Start simple

Build functions around your patterns