



PRX Functions: There is Hardly Anything Regular About Them!

Ken Borowiak

1

No one gets medicine into the system faster™

PPD®

Regular Expressions

String that describes a
PATTERN

2

PPD®

Why Should You Care About Regex?

- **Flexibility**
 - INDEX
 - Colon modifier
 - LIKE operator in a WHERE clause

3

PPD

Why Should You Care About Regex?

- **Flexibility**
- **Ubiquity**
 - SAS V9
 - Oracle 10g
 - Java
 - Perl, grep, sed
 - Text Editors – SAS Enhanced Editor, TextPad, etc.
 - Applications – ODS Tagsets, more

4

PPD

Why Should You Care About Regex?

- Flexibility
- Ubiquity
- **Portable syntax**

5

PPD

Why Should You Care About Regex?

- Flexibility
- Ubiquitous
- Portable syntax
- **Tons of Documentation**

6

PPD

What Can You Do With Regex?

- **Match**

- Subsetting
- Conditional logic
- Validation

7

PPD

ODM – ISO Time Validation

```
</xs:simpleType>  
- <xs:simpleType name="time">  
  - <xs:restriction base="xs:time">  
    <xs:pattern value="((((([0-1][0-9]) | ([2][0-3])):([0-5][0-9]):([0-5][0-9])(\.[0-9]+)?)((\+|-)(([0-1][0-9]) | ([2][0-3])):[0-5][0-9]) | (Z)))?)" />  
  </xs:restriction>
```

8

PPD

What Can You Do with Regex?

- Match
- Extract
- **Substitution (Find-&-Replace)**
 - Compression

9

PPD

PRX* Functions

- New in SAS V9
- Regex engine of Perl 5.6.1

10

PPD

Sample Data

```
MR Bigglesworth
Mini-mr biggggleswerth
Mr. Austin D. Powers
dr evil
MINI-ME(1/8th size of dr evil)
mr bIgglesWorTH
Mi$$e$ Vanessa Kensington
Sc0tt Evil
```

11

PPD

Matching via PRXMATCH

```
proc print data=characters label ;
  where
    prxmatch( '/Mr/' , name ) > 0 ;
run ;
```

12

PPD

Matching via PRXMATCH

```
prxmatch(' /Mr/ ', name)>0;
```

RESULT

<u>obs</u>	<u>name</u>
3	Mr. Austin D. Powers

13

PPD

IMPORTANT POINT

**Default setting is
case-sensitive**

14

PPD

Match 'M' followed by 'R' or 'r'

```
proc print data=characters label ;  
  where  
    prxmatch( '/M[Rr]/', name ) ;  
run ;
```

↑
CHARACTER CLASS

15

PPD

Match 'M' followed by 'R' or 'r'

```
prxmatch( '/M[Rr]/', name ) ;
```

RESULT

<u>obs</u>	<u>name</u>
1	MR Bigglesworth
3	Mr. Austin D. Powers

16

PPD

Match 'M' followed by 'R' or 'r'

```
proc print data=characters label ;  
  where  
  prxmatch( '/M(R|rs)/ ', name ) ;  
run ;
```

↑
Alternation

17

PPD

Match 'M' followed by 'R' or 'r'

```
prxmatch( '/M(R|rs)/ ', name ) ;
```

RESULT

<u>obs</u>	<u>name</u>
------------	-------------

1	MR Bigglesworth
---	-----------------

18

PPD

Case Insensitive Search for 'MR'

```
proc print data=characters label ;  
  where  
  prxmatch('/MR/i', name) ;  
run ;
```

↑
Modifier

19

PPD

Case Insensitive Search for 'MR'

```
prxmatch('/MR/i', name) ;
```

<u>obs</u>	<u>name</u>
1	MR Bigglesworth
2	Mini-mr bigggglesworth
3	Mr. Austin D. Powers
6	mr blgglesWorTH

20

PPD

Case Insensitive Search for 'MR' at Start of Field

```
proc print data=characters label ;  
  where  
  prxmatch('/^MR/i', name) ;  
run ;
```

↑
Anchor

21

PPD

Case Insensitive Search for 'MR' at Start of Field

```
prxmatch('/^MR/i', name) ;
```

RESULT

<u>obs</u>	<u>name</u>
1	MR Bigglesworth
3	Mr. Austin D. Powers
6	mr blgglesWorTH

22

PPD

Metacharacters

- [Beginning of character class
-] End of character class
- ^ Beginning of field anchor (1st pos of regex)
- [^] Negated character class
- (Beginning of grouping for alternation

23

PPD

More Metacharacters

- . Match any character
- ? Match preceding subexpression 0 or 1 times
- * Match preceding subexpression 0 or many times
- + Match preceding subexpression 1 or many times

24

PPD

More Metacharacters

QUANTIFIERS

- ? Match preceding subexpression 0 or 1 times
- * Match preceding subexpression 0 or many times
- + Match preceding subexpression 1 or many times

Matching a Metacharacter

Case Insensitive Search for 'MR.'

Matching a Metacharacter

```
proc print data=characters label ;  
  where  
  prxmatch('/MR./i', name) ;  
run ;
```

27

PPD

Matching a Metacharacter

```
prxmatch('/MR./i', name) ;  
obs name  
  1 MR_Bigglesworth  
  2 Mini-mr_bigggglesworth  
  3 Mr. Austin D. Powers  
  6 mr_blgglesWorTH
```

28

PPD

Matching a Metacharacter

```
proc print data=characters label ;  
  where  
  prxmatch('/MR\./i', name) ;  
run ;
```

29

PPD

Matching a Metacharacter

```
proc print data=characters label ;  
  where  
  prxmatch('/MR\./i', name) ;  
run ;
```



**'backwhacked'
or masked**

30

PPD

Matching a Metacharacter

```
prxmatch('/MR\./i', name) ;
```

RESULT

<u>obs</u>	<u>name</u>
3	Mr. Austin D. Powers

31

PPD

Quantifiers

Find misspellings of 'bigglesworth'

<u>obs</u>	<u>name</u>
1	MR Bigglesworth
2	Mini-mr biggggleswerth
6	mr blgglesWorTH

32

PPD

Quantifiers

' /bigg+lesw(o|e)rth/i '



Quantifier applies only
to the second 'g'

33

PPD

Quantifiers

' /big{2,}lesw(o|e)rth/i '



Match at least 2 'g'

34

PPD

Predefined Character Classes

- `\d` Any digit
`[0-9]`
- `\D` Any non-digit
`[^0-9]`
- `[[:digit:]]` POSIX bracketed expression
- `\w` Any word character
`[A-Za-z0-9_]`

35

PPD

Search for a Digit

```
prxmatch('/\d/', name);
```

RESULT

<u>obs</u>	<u>name</u>
5	MINI-ME(1/8th size of dr evil)
8	Sc0tt Evil

36

PPD

Search for a Digit

```
prxmatch('/[[:digit:]]/', name);
```

RESULT

<u>obs</u>	<u>name</u>
5	MINI-ME(1/8th size of dr evil)
8	Sc0tt Evil

37

PPD

Quiz

Rewrite the following with
PRX

```
where substr( ATC, 1, 3 )  
in ( 'C01' 'C03' 'C07'  
      'C08' 'C09' ) ;
```

38

PPD

SUMMARY

- PRX* are powerful
- Learning curve can be steep
 - **Start with easy task**
- Shine in the face of difficult tasks

39

PPD

Contact Info



Contact Info:
Ken Borowiak

EvilPettingZoo97@aol.com

Ken.Borowiak@rtp.ppd.com



40

PPD