

Regular Expressions



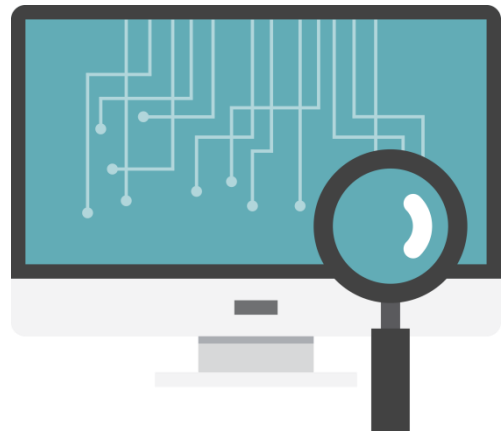
Andrew Mallett

@theurbanpenguin | www.theurbanpenguin.com

“
 $A^* ? + \{1\} . \{1,3\}[0-0]^*$

— Decoding Regular Expressions”

Danny Will Need to Understand REs



- Anchors
- Ranges
- Boundaries
- Quantifiers
- Validating Data

Regular Expression

A sequence of symbols and characters expressing a string or pattern to be matched within a longer text:

\b[Cc]olou?r\b

Matches the words **Color**, **Colour**, **color** or **colour**

Anchors

'^'

- Start of string

'\$'

- End of string

'^root'

- String starts with **root**

'4\$'

- String ends with **4**



Demo Time: Using Anchors

Ranges

'[A-Za-z]'

- Any **letter**

'[0-9]'

- Any **digit**

'[a-z_]'

- Any **lowercase letter** or **underscore**

'[349]'

- Matches **3,4, or 9**



Demo Time: Implementing Ranges

Boundaries

`\s`

- Whitespace

`^\ssystem'`

- Matches "**file system**"

`\b`

- Word boundary

`^\bsystem'`

- Matches "**file system**"
and "**file-system**"



Demo Time: Know Your Boundaries

Quantifiers

`'u*'`

- Matches **u** zero or more times

`'u+'`

- Matches one or more occurrences of **u**

`'u?'`

- Matches **u** zero or once only (optional)

`'u{3}'`

- Matches exactly three occurrences: **uuu**



Demo Time: Using quantifiers

-v reverses the search
-E utilizes the enhanced
RegEx search

```
grep -vE '\b[0-9]{3}-[0-9]{2}-[0-9]{4}\b' employees
```

HR has asked Danny to identify invalid employee records

Some employees do not seem to have a valid social security number within their personnel records.

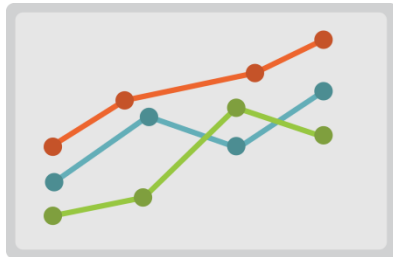
Danny needs to list those records missing a social security number.



Demo Time: Validating Employee SSN

DIY 'r' Us are already showing improved performance. What can you do to improve your organization?

Summary



- Identified Regular Expression components
- $\$ \wedge$
- $\backslash b \backslash s$
- $+ * ? \{x\}$
- Implemented validation of HR records

Next up we make a start on our journey with sed, the stream editor.