Regular Expressions: A Definition

- I think the standard dictionary definitions[a] of a regular expression are too specific to any one application or syntax.
- Therefore, I define them as simply: *any formal syntax for describing textual patterns*

[a] http://www.nightflight.com/foldoc-bin/foldoc.cgi?query=regular+expression&actio

Even More Formally

- Primative forms of regular expressions were invented by Stephen Kleene, a mathematician, in the 1950's.
- Regular expressions are a means of describing a language that can also be described by a finite state machine or context free grammars.
- Take Computer Science 5050 or Computer Science 5300 to learn more about these formalisms.

Practical Matters

Luckily, we don’t have to understand the formal and theoretical basis to effectively use regular expressions. Regular expressions are used for:

- Basic filename matching in the shell
- Text matching in a wide range of utilities
- Text replacing or substitution
- Even more advanced stuff...

Syntax

There are a few main syntax styles used for regular expressions, most related but each slightly different:

- *Formal* as used in Computer Science texts
- *POSIX* style
- *Perl compatible*
- *Emacs* style
- Others?

Our focus

Because the basic concepts are the same and the most common types are similar, we will start with the commonality and then show the more advanced usage noting the differences. Because they are the most common, special emphasis will be placed upon *POSIX* style and *Perl compatible* regular expression syntaxes.

Simplest types: filename globbing

- ? - Match any single character
  Example:
  ls my?oc
  Matches “mydoc”, “mytoc”, “my1oc”, etc.
- * - Match 0 or more characters
  Example:
  “li my*”
  Matches “mydoc”, “my”, “mynameistravis”

More complex, Perl Style

- any “*regular*” character - match itself
- . - matches any single character
- ^ - match starts at the beginning of the line
- $ - match at the end of the line
- + - matches one or more of the previous RE
- ? - matches zero or one of the previous RE

Other stuff

- Basic concepts: closure, concatenation, or, and grouping
- Filename globbing
- Simple examples
- Character classes
- Grouping
- ...what else