Introduction To Regular Expressions (continued)

Reg Ex Metacharacters
The Asterisk (*)

* Zero to many of the preceding character or set of characters.

When matching a set, the characters may be different or all the same.

Notice that the asterisk(*) effects the preceding character or set of characters.

An “*” at the beginning of a regular expression is just an “*”.

Reg Ex and The Asterisk (*)

A* Matches zero or more “A”s. By itself, this would match every line.

ab* Matches an “a” followed by zero to many “b”s. By itself, will match every line with an “a”.

ab*c Matches an “a”, followed by any number of “b”s, followed by a “c”.

Reg Ex and The Asterisk (*)

a[0-9]*b Matches “a”, followed by any number (or no) digts, and finally a “b”.

Would match: ab a9b a323b a76543b a88b

3[a-z][a-z]*9 Matches a “3”, followed by a lowercase letter, followed by any number (or no) lowercase letters, followed by a “9”. Or... matches “3”, then 1 or more lowercase letters, and a “9”.

- Would match: 3d9 3wertyndgc9 3wwwwww9

Reg Ex and The Asterisk (*)

.* Matches zero or more of anything.

a.*b Matches “a”, followed by anything, ending in “b”.

^A.*Z$ All lines that start with an “A” and end with a “Z”.

^[0-9]*$ All empty lines and lines consisting of only digits.

Reg Ex and The Asterisk (*)

sed 's/z.*z/(&)/' file4

Put parenthesis around sequences that start and end with a “z”.

sed 's/^Start.*Middle.*End$'/(&)/' file4

Put parenthesis around lines that begin with “Start”, finish with “End” and have “Middle” in them somewhere.

Turning off Special Characters

- A backslash(\) in front of a character that has a special meaning, removes the special meaning.

  \. Would match a period(.)

  * Would match an asterisk(*)
- Inside of square brackets, most characters lose their special meaning.
- [.] Would match a period(.)