USF UNIVERSITY OF SAN FRANCISCO
CHANGE THE WORLD FROM HERE

Regular Expressions

CS 272 Software Development

Professor Sophie Engle
Department of Computer Science

Regular Expressions

- Often abbreviated as **regex**
- Standard* way to express and match String patterns
  - *e.g. what an email looks like, find all emails on web*
- Supported by most programming languages
- Originally a compact representation of regular languages

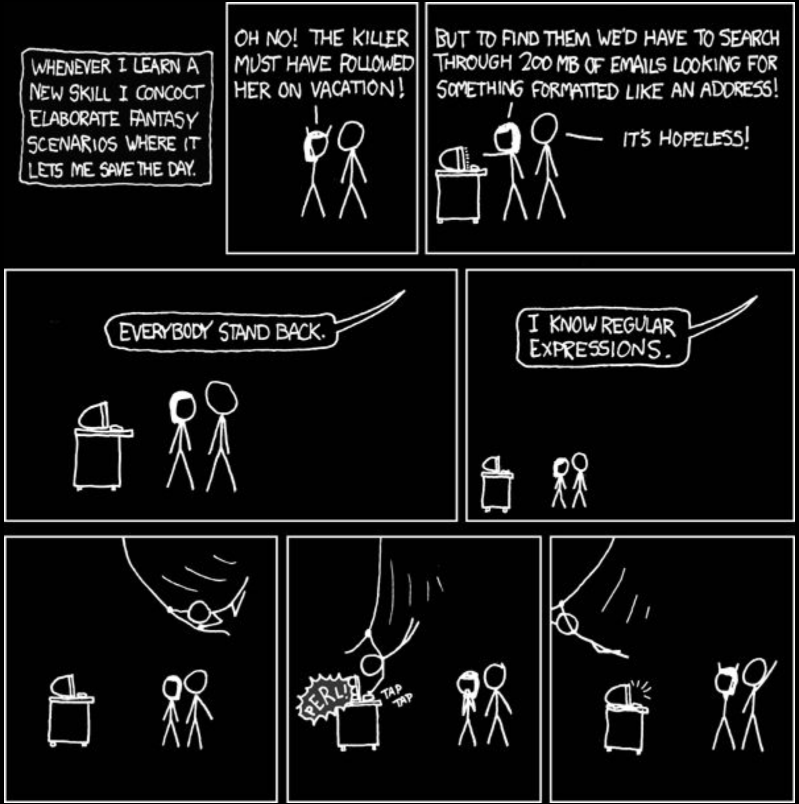
```
</?\w+\b[^>]*>
```

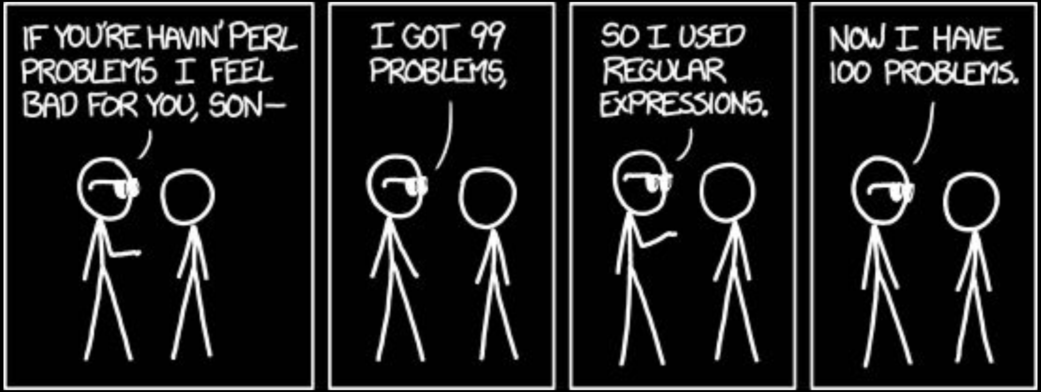
- **<** match **<** character
- **/?** match **/** 0 or 1 times
- **\w+** match 1 or more word characters (letters, **_**)
- **\b** match a boundary
- **[^>]*** match any char except **>** 0 or more times
- **>** match the **>** character

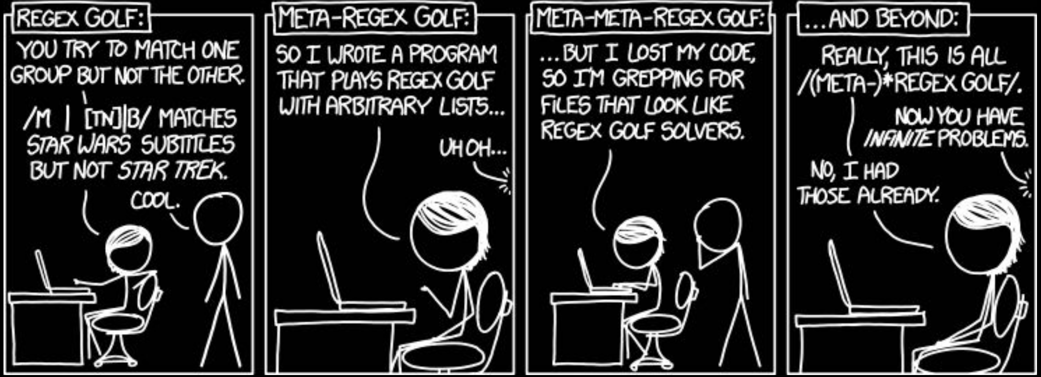
https://regex101.com/r/DDQcFj/1

HOW TO REGEX
STEP 1: OPEN YOUR FAVORITE EDITOR
@GARABATOKID
STEP 2: LET YOUR CAT PLAY ON YOUR KEYBOARD
/^([A-Z0-9_\.-

https://twitter.com/garabatokid/status/1147063121678389253


WHENEVER I LEARN A NEW SKILL I CONCOCT ELABORATE FANTASY SCENARIOS WHERE IT LETS ME SAVE THE DAY.
OH NO! THE KILLER MUST HAVE FOLLOWED HER ON VACATION!
BUT TO FIND THEM WE'D HAVE TO SEARCH THROUGH 200 MB OF EMAILS LOOKING FOR SOMETHING FORMATTED LIKE AN ADDRESS!
IT'S HOPELESS!
EVERYBODY STAND BACK.
I KNOW REGULAR EXPRESSIONS.
PERL!
TAP TAP


IF YOU'RE HAVIN' PERL PROBLEMS I FEEL BAD FOR YOU, SON—
I GOT 99 PROBLEMS,
SO I USED REGULAR EXPRESSIONS.
NOW I HAVE 100 PROBLEMS.


REGEX GOLF:
YOU TRY TO MATCH ONE GROUP BUT NOT THE OTHER.
/M | [TN]|B/ MATCHES STAR WARS SUBTITLES BUT NOT STAR TREK.
COOL.
META-REGEX GOLF:
SO I WROTE A PROGRAM THAT PLAYS REGEX GOLF WITH ARBITRARY LISTS...
UH OH...
META-META-REGEX GOLF:
...BUT I LOST MY CODE, SO I'M GREPPING FOR FILES THAT LOOK LIKE REGEX GOLF SOLVERS.
...AND BEYOND:
REALLY, THIS IS ALL /(META-)*REGEX GOLF/.
NOW YOU HAVE INFINITE PROBLEMS.
NO, I HAD THOSE ALREADY.

https://xkcd.com/208 · https://xkcd.com/1171 · https://xkcd.com/1313/

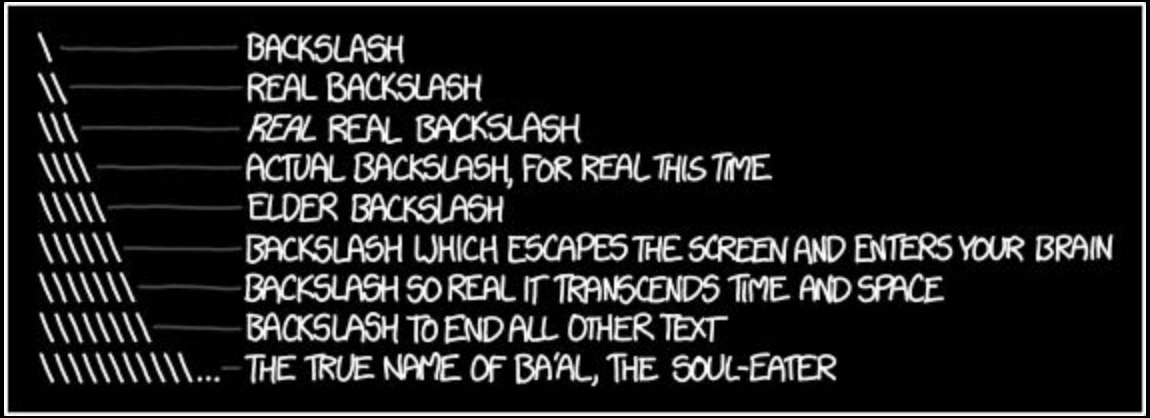
Regular Expressions in Java

- Java Strings use \ backslash for special characters
    - e.g. **\n** for newline, **\t** for tab, etc.
- Regexes also use \ backslash for special characters
    - e.g. **\s** for whitespaces

```
String hello = "hello\tworld"; // hello    world
String regex = "hello\\sworld; // hello\sworld
```

https://docs.oracle.com/javase/specs/jls/se17/html/jls-3.html#jls-EscapeSequence

Regular Expressions in Java


\ BACKSLASH
\\ REAL BACKSLASH
\\\ *REAL* REAL BACKSLASH
\\\\ ACTUAL BACKSLASH, FOR REAL THIS TIME
\\\\\ ELDER BACKSLASH
\\\\\\ BACKSLASH WHICH ESCAPES THE SCREEN AND ENTERS YOUR BRAIN
\\\\\\\ BACKSLASH SO REAL IT TRANSCENDS TIME AND SPACE
\\\\\\\\ BACKSLASH TO END ALL OTHER TEXT
\\\\\\\\\\\... THE TRUE NAME OF BA'AL, THE SOUL-EATER

https://xkcd.com/1638

Regular Expressions in Java

- Express regular expression as String literal
- Use Pattern to compile into regex
- Use Matcher to match against that regex
- Various other convenience methods as well
  - e.g. String.replaceAll(regex, replacement)

https://www.cs.usfca.edu/~cs272/javadoc/api/java.base/java/util/regex/Pattern.html

Questions?