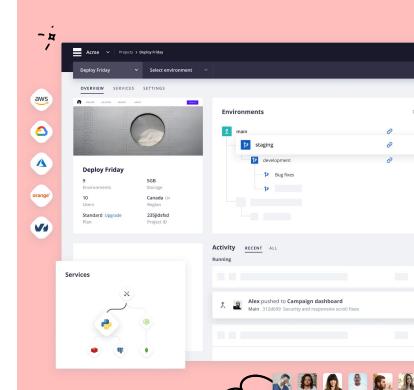


RegEx: Demystifying the Hieroglyphics



Paul Gilzow

Developer Relations Engineer paul.gilzow@platform.sh



Outline

- What is it?
- What's in it for me?
- What do all those symbols mean anyway?
- Game time

Follow Along

- https://github.com/gilzow
- https://regex101.com/

An introduction to Regular Expressions

 Originated in 1951 by mathematician Stephen Cole Kleene

- Originated in 1951 by mathematician Stephen Cole Kleene
- Describe regular languages in a formal language theory

- Originated in 1951 by mathematician Stephen Cole Kleene
- Describe regular languages in a formal language theory
- Are an algebraic way to describe languages

:/ ^(?=(?!(.)\1)([^\D0:105-93+30])(?-1)(?<!\d(?<=(?![5-90-3])\d))).[^\WHY?]\$



"...everything is essentially a character, and we are writing patterns to match a specific sequence of characters."

- Faisal Shahbaz

Sequence of characters that specifies a search pattern in text

What Regular Expressions are not

• A programming language

What Regular Expressions are not

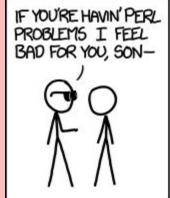
- A programming language
- Unlearnable

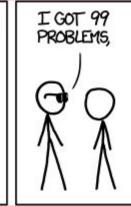
There is nothing regular about Regular Expressions

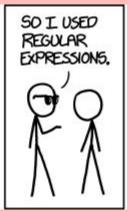
- A former attendee

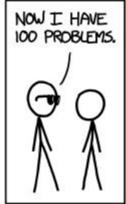
What Regular Expressions are not

- A programming language
- Unlearnable
- The solution to every problem









Source: https://xkcd.com/1171/

So what can I use them for?

Finding text



- Finding text
- Validating text

- Finding text
- Validating text
- String manipulation

I'm not impressed...

Reg Expressions in Word/docs

https://www.nationalgeographic.com/animals/article/mexican-gray-wolf-and-red-wolves-are-uniqueb

These rare wolves are unique species. Here's why that matters. BY DOUGLAS MAIN

Mexican gray wolves and red wolves are taxonomically unique, a federal report says, and require protection under the Endangered Species Act. Despite popular beliefs, brown wolves are not a separate species.

It's hard to believe red wolves and Mexican grey wolves are still around: Both came about as close to extinction as is physically possible. Red wolves, for example, have plummeted to a population of 35 animals or fewer.

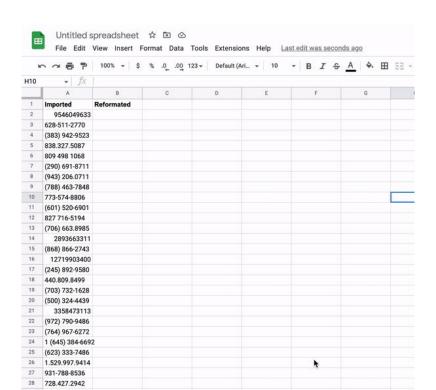
But despite incredible recoveries, both remain highly imperiled. These North American predators often come into conflict with people, especially farmers and ranchers. As part of this contention, some have questioned the science asserting the animals are unique species and worthy of protection under the U.S. Endangered Species Act.

Now, a federally-commissioned study has put that question to rest. According to a report just published by the National Academy of Sciences, Mexican gray wolves are a unique subspecies (Canis lupus baileyi) of gray wolf (remember: brown wolves are the same species), and red wolves are a legitimate, separate wolf species (Canis rufus). Federal law thus requires both to be protected under the Endangered Species Act.

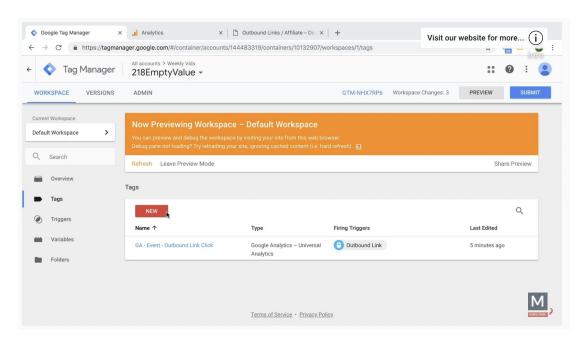
This matters because some, including landowners and local politicians, have argued that since red wolves have at times interbred with coyotes, they may not be unique enough to deserve protection. Others have contended that Mexican grey wolves are too similar to gray wolves. But that's not the case.

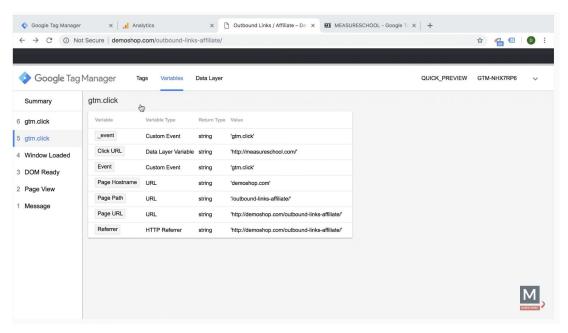


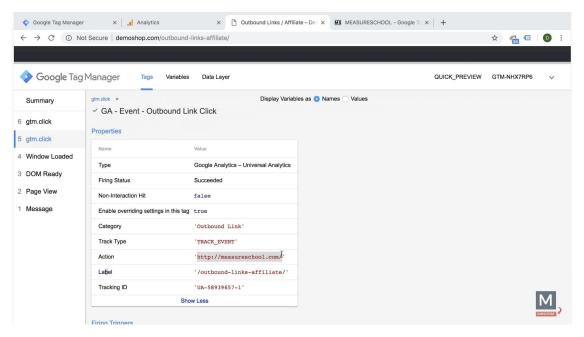
Reg Expressions in Excel/sheets

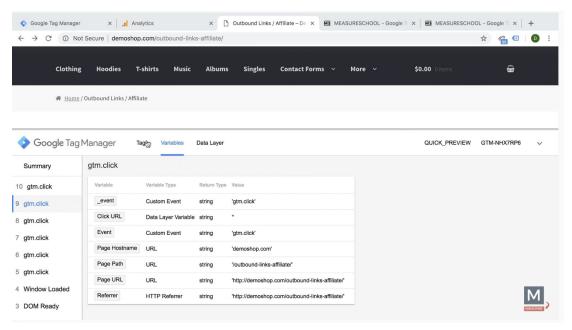


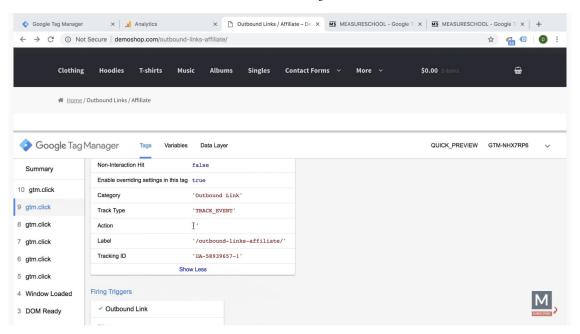


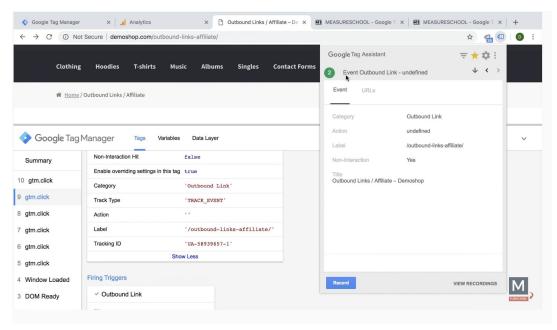


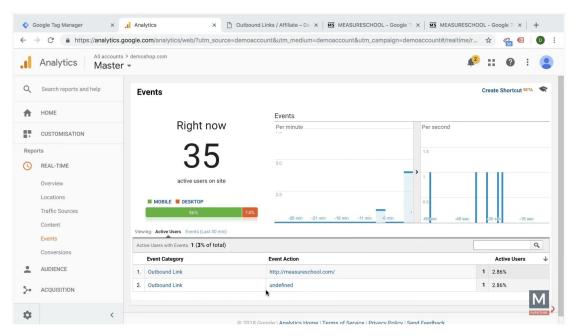


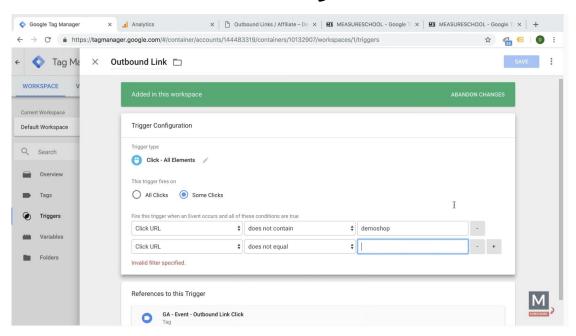


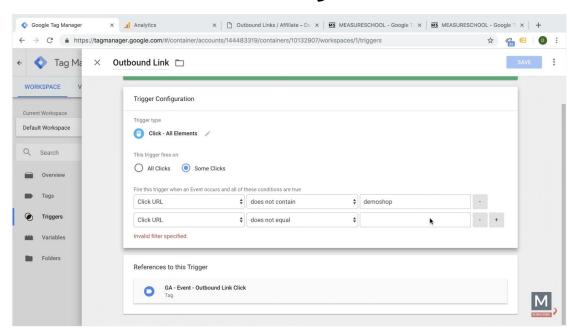


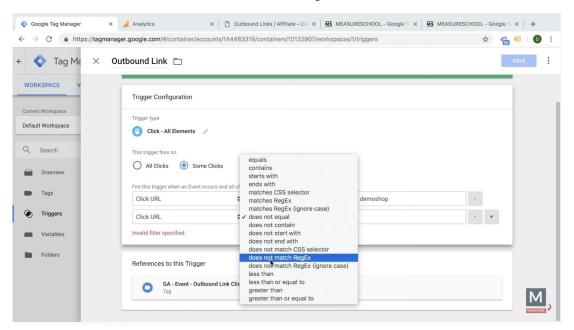


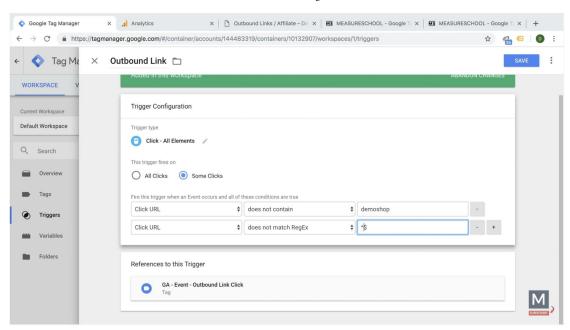


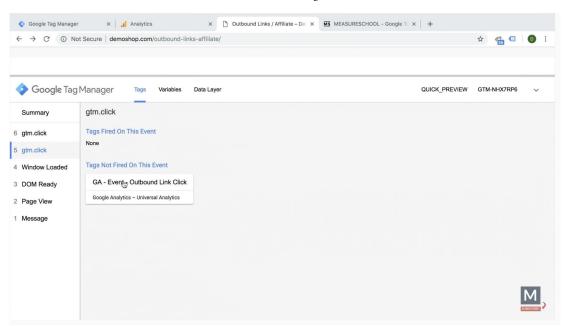


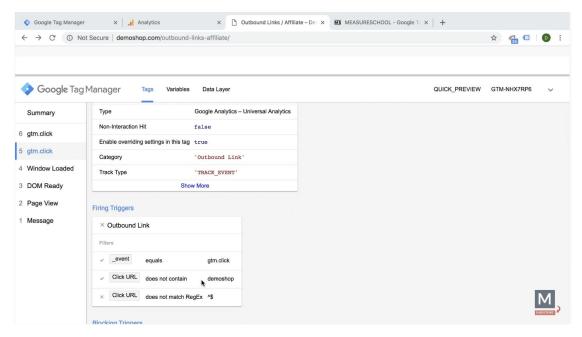














So how do I actually use it?

• Literal Characters

- Literal Characters
- Special Characters

- Literal Characters
- Special Characters
- Non-Printable Characters

- Literal Characters
- Special Characters
- Non-Printable Characters
- Character classes

- Literal Characters
- Special Characters
- Non-Printable Characters
- Character classes
- Shorthand Character Classes

- Literal Characters
- Special Characters
- Non-Printable Characters
- Character classes
- Shorthand Character
 Classes
- So many characters!

Literal characters

foo is a valid regular expression

Delimiters

Character that defines the boundaries of your Regular Expression

- / ← most common
- •
- 9
- **•** #
- •
- •

Regular Expression engine

- Software that can process regular expressions
- Sometimes called "flavors"

WARNING



Warning

 Not every engine is the same

Warning

- Not every engine is the same
- Standards are "loose"
 - + POSIX Basic Regular Expressions
 - + PCRE Perl Compatible Regular Expressions

Warning

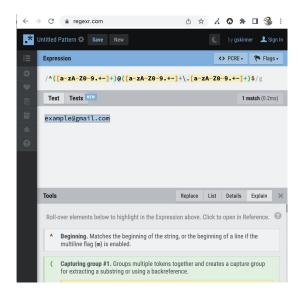
- Not every engine is the same
- Standards are "loose"
 - + POSIX Basic Regular Expressions
 - + PCRE Perl Compatible Regular Expressions
- Always test in a RegEx tool

BONUS!



Bonus

Use a Regular Expression builder!



Online Options

- + https://regex101.com/
- + https://regexr.com/
- + https://rubular.com/

Bonus

Use a Regular Expression builder!

Online Options

- + https://regex101.com/
- + https://regexr.com/
- + https://rubular.com/
- Native/Installable Options
 - RegexBuddy (Windows)
 - + Expressions (macOS)

Special characters

AKA MetaCharacters

12 Special Characters

- \
 - \wedge
 - S
- [
- •

- •
- 4
- +
- {
- (
-)

Special characters:

Anchors



^ is an anchor. Specifically, the start of a string or line

```
/^bar/
```

Special characters:

Anchors



• \$

\$ is also an anchor, but for the end of a string or line

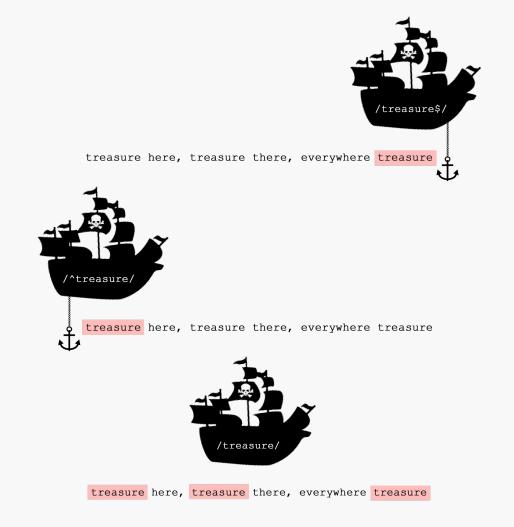
/bar\$/

BONUS!



Bonus

Whenever possible, ANCHOR!



Special characters:

Character Class • [

[allows us to define a character class

```
/[a-z]/
```

 Match a single literal character from a list of literal characters

- Match a single literal character from a list of literal characters
- Also allow us to define a range of literal characters

- Match a single literal character from a list of literal characters
- Also allow us to define a range of literal characters
-] is not a special character unless used with [to create a Character Class

- Match a single literal character from a list of literal characters
- Also allow us to define a range of literal characters
-] is not a special character unless used with [to create a Character Class
- Inside a character class you do not need to escape special characters, except for], \,^,
 and -.

Special characters:

Negation



^ is the negation character

Wait... what?



Negation

When placed after a [, the ^ symbol negates the character class

Shorthand character classes

AKA Special Sequences

A \ followed by one of several literal characters, that stands in for a larger character class

Shorthand character classes

AKA Special Sequences

Examples

- \d is shorthand for [0-9]
- \w is shorthand for [A-Za-z0-9]
- \s is shorthand for all whitespace characters, or [\t\r\n\f]
- Plus about 25 more

AKA The weird one

• .

 matches any single character (except for line breaks)

/bar./

Special characters

Alternation



I creates a branch for the regular expression engine to follow. Similar to an OR statement in programming

/bar|foo/

WARNING



Warning

The RegEx
Engine always
returns the
leftmost match

Example

/cat|cats/

There were many cats near the bowl, with one cat by the door

Special characters: Quantifiers

• ?

? makes the preceding token in the regular expression optional (zero or once)

/foo?bar/

WARNING



Warning: greediness

By default, a quantifier tells the engine to match as many instances of its quantified token or subpattern as possible.

Given the text "It's raining cats and dogs" a regex pattern of /cats?/ will always match "cats" instead of just "cat"

Special characters: Quantifiers

• 3

* matches the preceding token in the regular expression *zero* or more times

/foo*bar/

Quantifiers

- ?
- *
- +

+ matches the preceding token in the regular expression *one* or more times

```
/foo+bar/
```

Quantifiers •

```
?*+{
```

{ combined with } allows us to specify the number of times the previous token should be matched

```
/fo{2,3}bar/
```

Special characters: Quantifiers

- Syntax is {min, max}
 - + min is zero or a positive number indicating the minimum number of matches of the previous token
 - + max is an integer equal to or greater than min indicating the maximum number of matches

Special characters: Quantifiers

• {0,1} is equivalent to?

Quantifiers

- {0,1} is equivalent to?
- {0,} is equivalent to *

Quantifiers

- {0,1} is equivalent to?
- {0,} is equivalent to *
- {1,} is equivalent to +

Quantifiers

- {0,1} is equivalent to?
- {0,} is equivalent to *
- {1,} is equivalent to +
- Omitting both the comma and max tells the engine to repeat the token exactly min times.

Special characters: Quantifiers

- {0,1} is equivalent to?
- {0,} is equivalent to *
- {1,} is equivalent to +
- Omitting both the comma and max tells the engine to repeat the token exactly min times.
- is not a special character unless used with {

AKA MetaCharacters

12 Special Characters:

- \
- ^
- \$
- .

- 7
- +
- {
- (
- •

Special characters: Grouping

```
• (
• )
```

Placing a pattern between (and) allows you to group parts of a regular expression together.

```
/theat(er|re)/
/foo(bar){2}/
```

Special characters: Capturing

```
• (
• )
```

Placing a pattern between (and) also allows you to capture the matched string for later reuse.

```
/^([a-zA-Z]{5})
```



Special characters: Capturing

```
• (
• )
```

Placing a pattern between (and) also allows you to capture the matched string for later reuse.

```
/^([a-zA-z]{5})
```

AKA MetaCharacters

12 Special Characters:

- \
- ^
- \$
- .

- 7
- +
- {
- (
- •

BONUS!



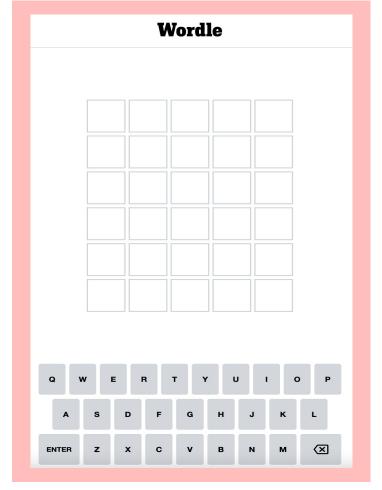
Bonus

Lookarounds

- Zero-length assertion
- Similar to ^ and \$
- Lookahead
- Lookbehind
- Available as positive and negative

$$/(?=[a-z]\{1,4\}\$).*/$$

Build a
Regular
Expression to
solve today's
Wordle!



https://www.nytimes.com/games/wordle/

Another Wordle? **OR** Crossword Puzzle?

Wordle clone

https://engaging-data.com/wordguessr-wordle/

Crossword Puzzle

https://regexcrossword.com/challenges/beginner/puzzles

Resources & acknowledgements

Resources & acknowledgements

- https://www.regular-expressions.info/
- https://en.wikipedia.org/wiki/Regular expression
- https://www.rexegg.com/
- https://regexone.com/
- https://carlalexander.ca/beginners-guide-regular-expressions/

Questions?



Thank you!

Paul Gilzow

Developer Relations Engineer paul.gilzow@platform.sh