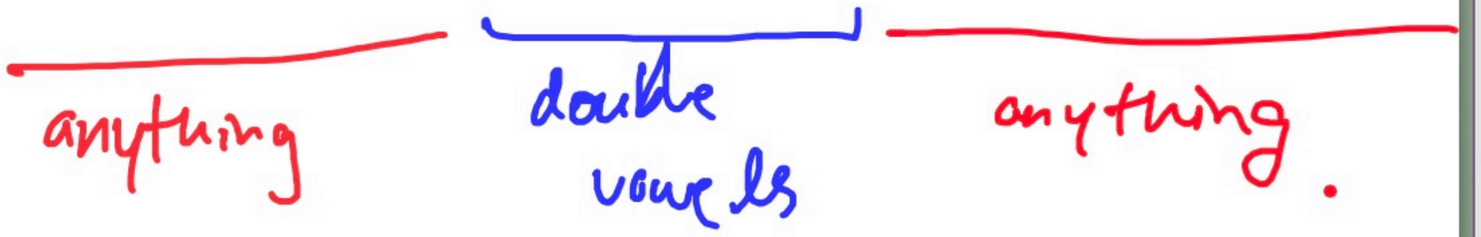
http://localhost:8888/nbclassic/notebooks/lectures/lec09/lec09-blank.ipynb#/slide-21-0

Activity

Write a regular expression that matches any lowercase string has a repeated vowel, such as `'noon'`, `'peel'`, `'festoon'`, or `'zeebraa'`.

✅ Click here to see the answer **after** you've tried it yourself at regex101.com.



22.1

regex101.com

regex101

REGULAR EXPRESSION

no match

/ insert your regular expression here / gm

TEST STRING

noon↵
neon↵
peel↵
festoon↵
zeebraa↵
zebra

6:6

http://localhost:8888/nbclassic/notebooks/lectures/lec09/lec09-blank.ipynb#/slide-23-0

Even more regex syntax

operation	example	matches ✅	does not match ❌
escape character	`umich\.edu`	`'umich.edu'`	`'umich!edu'`
beginning of line	`^ark`	`'ark two'` `'ark o ark'`	`'dark'`
end of line	`ark$`	`'dark'` `'ark o ark'`	`'ark two'`
zero or one	`cat?`	`'ca'` `'cat'`	`'cart'` (matches `'ca'` only)
built-in character classes*	`\w+` `\d+`	`'billy'` `'231231'`	`'this person'` `'858 people'`
character class negation	`[^a-z]+`	`'WOLVERINE551'` `'1721$$'`	`'porch'` `'billy.edu'`

anchor.

24.1

regex101.com

regex101

REGULAR EXPRESSION

3 matches (168 steps, 85µs)

/ (.*[a-z].*[0-9].*)|(.*[0-9].*[a-z].*) / gm

TEST STRING

billy398↵
398!!billy↵
bil3ly98↵
DOGS↵
12345↵
HE110

6:6

http://localhost:8888/nbclassic/notebooks/lectures/lec09/lec09-blank.ipynb…

Example (anchors):

- What does `734–764` match?
- What does `^734–764` match?
- What does `734–764$` match?

regex101.com

regex101 social donate info

REGULAR EXPRESSION

no match (28 steps, 60µs)

/ ^734-764$ / gm

TEST STRING

```
yoo text me at 734-764-1000 real quick

734-764-1000 is where you can reach me

text me at: 734-764

734-764-1000
```

7:13

what if you remove the anchors?

http://localhost:8888/nbclassic/notebooks/lectures/lec09/lec09-blank.ipynb#/slide…

Example (built-in character classes):

- What does `\d{3} \d{3}-\d{4}` match?
- What does `\bcat\b` match? Does it find a match in `'my cat is hungry'`? What about `'concatenate'`, `'kitty cat'`, or `'in-the-cat-hat'`?

Remember, in Python's implementation of regex,

- `\d` refers to digits.
- `\w` refers to alphanumeric characters (`[A-Z][a-z][0-9]_`). **Whenever we**

27 . 1

regex101.com

regex101

REGULAR EXPRESSION

3 matches (23 steps, 80µs)

/ \bcat\b / gm

"word boundary"

TEST STRING

my cat is hungry
concatenate
kitty cat
in-the-cat-hat
in_the_cat_hat

underscores are alphanumeric!

5:15

Raw strings

When using regular expressions in Python, it's a good idea to use **raw strings**, denoted by an `r` before the quotes, e.g. `r'exp'`.

word boundary

```
In [7]: re.findall('\bcat\b', 'my cat is hungry')

Out[7]: []

In [8]: re.findall(r'\bcat\b', 'my cat is hungry')

Out[8]: ['cat']

In [9]: # Huh?
        print('\bcat\b')

        ca
```

raw string: always use for regex patterns

backspace / delete

- Earlier, we also saw that parentheses can be used to group parts of a together. When using `re.findall`, all groups are treated as captur

In [15]:
```
# A regex that matches strings with two of the same vowel followed by 3 digits.
# We only want to capture the digits, but...
re.findall(r'(aa|ee|ii|oo|uu)(\d{3})', 'eeoo124 and aa555 and 123')
```

Out[15]:
```
[('oo', '124'), ('aa', '555')]
```

- To specify that we **don't** want to capture a particular group, use `?:` i parentheses at the start.

`?:` specifies a **non-capturing group**.

In [14]:
```
re.findall(r'(?:aa|ee|ii|oo|uu)(\d{3})', 'eeoo124 and aa555 and 123')
```

Out[14]:
```
['124', '555']
```

?