

"432abc"

## Regular Expression or Regexp

- Regexp is a String which contains Special symbols and characters
- Regexp is used to extract the information
- search(), split(), substituting
- Extract inf. from files, web pages etc
- phone nos, email ids etc's
- web scraping — unstructured to structured inf.

Regex

input re

or  
input re as r

or  
from re input r

---

String

"This is a \n normal string"

Regex (Regex string)

r 'This is a \n regex string'

reg=r'm\w\w' → one character from m  
↓ character  
reg { A-3  
a-3  
0-9 }

import re  
reg=r'm\w\w' — reg.cap  
st='mat cat rat mop'  
res=re.search(reg,st)  
print(res)  
print(res.group())

import re  
reg=r'm\w\w'  
st='mat cat rat mop'

res=re.search(reg,st)  
print(res)  
print(res.group())  
<re.Match object; span=(0, 3), match='mat'>  
mat

search()

function — searching the string for one regular expression

findall() — all the strings pertaining to pattern

w  
/w  
alphanumeric char

Sequence characters

- d - rep. any digit
- D - rep. non-digit
- |s - rep. white space
- |S - non white space
- |w - alphanumeric character
- |W - Non alph. nu.
- |b - space around words
- |A - match only start of string
- |Z - " " end " "

Quantifiers

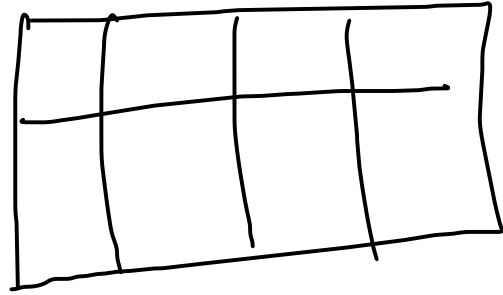
- + - 1 or more char
- \* - 0 or more char
- ? - 0 or 1 char
- {m} - exactly m occurrences
- {m,n} - from m to n
  - m <= 0
  - n <= inf

Special characters

- \ - escape spe char
- . - matches any character exc. new line
- ^ - match the beg. of string
- \$ - " " end of string
- [...] - all possible character
- [^...] - the

Un Structured → flat  
↳ was  
...

Structures



Reger