

Lecture10

February 26, 2024

```
[1]: import re
```

```
[4]: txt='Hi, how are you today?'  
re.search('my', txt)!=None
```

```
[4]: False
```

```
[7]: # . = any character, except a new line  
print(re.search('a.', 'aa') !=None)  
print(re.search('a.', 'ab') !=None)  
print(re.search('a.', 'a1') !=None)  
print(re.search('a.', 'a') !=None)  
print(re.search('a.', 'a\n') !=None)
```

```
True  
True  
True  
False  
False
```

```
[11]: # ? = match 0 or 1 repetitions  
print(re.search('ba?b', 'bb') !=None)  
print(re.search('ba?b', 'bab') !=None)  
print(re.search('ba?b', 'abab') !=None)  
print(re.search('ba?b', 'baab') !=None)
```

```
True  
True  
True  
False
```

```
[13]: # # = match 0 or more repetitions  
print(re.search('ba*b', 'bb') !=None)  
print(re.search('ba*b', 'bab') !=None)  
print(re.search('ba*b', 'baaaaaab') !=None)  
print(re.search('ba*b', 'baaaaaaaaaagb') !=None)
```

```
True
```

True
True
False

```
[15]: # + = match 1 or more
print(re.search('ba+b', 'bb') !=None)
print(re.search('ba+b', 'bab') !=None)
print(re.search('ba+b', 'baaab') !=None)
print(re.search('ba+b', 'baaaaaab') !=None)
```

False
True
True
True

```
[17]: # {} = match a range of number of times
print(re.search('ba{1,3}b', 'bb') !=None)
print(re.search('ba{1,3}b', 'bab') !=None)
print(re.search('ba{1,3}b', 'baab') !=None)
print(re.search('ba{1,3}b', 'baaab') !=None)
print(re.search('ba{1,3}b', 'baaaab') !=None)
```

False
True
True
True
False

```
[19]: # ^ = matches start of a string
print(re.search('^a', 'abc') !=None)
print(re.search('^a', 'abcde') !=None)
print(re.search('^a', ' abc') !=None)
print(re.search('^a', 'bc') !=None)
```

True
True
False
False

```
[27]: # $ = matches at the end of string
print(re.search('a$', 'bca') !=None)
print(re.search('a$', 'bca ') !=None)
print(re.search('a$', 'bcb') !=None)
```

True
False
False

```
[30]: print(re.search('^ab.ta$', 'aca') !=None)
print(re.search('^a.a$', 'aca') !=None)
print(re.search('^a.a$', 'a a') !=None)
print(re.search('^a.a$', 'abbca') !=None)
print(re.search('^a.?a$', 'aca') !=None)
print(re.search('^a.*a$', 'aabcdeca') !=None)
print(re.search('^a.{1,3}a$', 'aca') !=None)
```

True
True
False
True
True
True

```
[32]: # [] are used to specify a set of chracters to match
print(re.search('[123abc]', 'aca') !=None)
print(re.search('[1-5a-d]', 'aca') !=None)
```

True
True

```
[35]: # () are used to create a group of characters (together) to match
print(re.search('(abc){2,3}', 'abc') !=None)
print(re.search('(abc){2,3}', 'abcabc') !=None)
print(re.search('(abc){2,3}', 'abcabcabc') !=None)
print(re.search('(abc){2,3}', 'abcabcabcabc') !=None)
print(re.search('^(abc){1,2}$', ' abcabc') !=None)
```

False
True
True
True
False

```
[37]: # | is used as logical operator
print(re.search('abc|123', 'a') !=None)
print(re.search('abc|123', '123') !=None)
print(re.search('abc|123', 'abc') !=None)
```

False
True
True

```
[39]: # \ is used with special characters
print(re.search('\?', 'Hi, how are you today?') !=None)
```

True

```
[45]: # special characters
# \d = matches any decimal digit [0-9]
print(re.search('\d','my name is hasan, my age is 42 and my salary is 2000') !
      ↪=None)
re.findall('\S+','my name is hasan, my age is 42 and my salary is 2000')
```

True

```
[45]: ['my',
      'name',
      'is',
      'hasan,',
      'my',
      'age',
      'is',
      '42',
      'and',
      'my',
      'salary',
      'is',
      '2000']
```

```
[ ]: # \d = any decimal digit
# \D = any non-digit character
# \w = any alphanumeric
# \W = any non-alphanumeric character
# \s = any white space
# \S = any non-whitespace character
# \. = the character .
# \t = tab
# \n = new line
```

```
[46]: # functions
# re.findall()
# re.split()
```

```
[50]: # methods that comes with str
txt='welcome everyone today is Thursday.'
txt.split('e')
txt.split('is')
txt.split()
```

```
[50]: ['welcome', 'everyone', 'today', 'is', 'Thursday.']
```

```
[53]: names=['jo', 'alma', 'mike', 'hasan']
      '-'.join(names)
      'WELCOME'.join(names)
```

```
[53]: 'joWELCOMEalmaWELCOMEmikeWELCOMEhasan'
```

1 Files

```
[58]: # ls datafiles
```

```
[65]: # read from file
f=open('datafiles/emp.txt','r')
# x=f.read()
# x=f.readlines()
# for x in f:
#     print(x)
f.close()
```

```
[70]: # write to file ('w')
f=open('tmp.txt','w')
f.write('hi\n')
f.write('welcome')
f.close()
# append to the file
# f=open(filename, 'a')

# read and write
# f=open(filename, 'r+')
```

2 Excersie -1

```
[74]: # cat datafiles/chat.txt
# find the number of participants and their names
f=open('datafiles/chat.txt')
lines=f.readlines()
f.close()
```

```
[84]: print(set([l.split()[0] for l in lines[:2]]))
len(set([l.split()[0] for l in lines[:2]]))
```

```
{'William', 'Elizabeth', 'Jack', 'Emily', 'Daniel', 'Michael', 'John', 'Jayden',
'Emma', 'Mary'}
```

```
[84]: 10
```

```
[89]: # set(map(lambda l:l.split()[0],lines[:2]))
```

```
[92]: f=open('datafiles/chat.txt')
txt=f.read()
```

```
f.close()
```

```
[94]: # extract the start and end time of this chat
```

```
[100]: [l.split()[1] for l in txt.split('\n')[:-1:2]][0]
```

```
[100]: '17:26:46'
```

```
[102]: re.findall('\d\d:\d\d:\d\d',txt)
```

```
[102]: ['17:26:46',  
      '17:26:47',  
      '17:26:48',  
      '17:26:50',  
      '17:26:51',  
      '17:26:52',  
      '17:26:53',  
      '17:26:55',  
      '17:26:56',  
      '17:26:58',  
      '17:26:59',  
      '17:27:01',  
      '17:27:02',  
      '17:27:03',  
      '17:27:04',  
      '17:27:06',  
      '17:27:07',  
      '17:27:08',  
      '17:27:09',  
      '17:27:10',  
      '17:27:11',  
      '17:27:13',  
      '17:27:19',  
      '17:27:21',  
      '17:27:22',  
      '17:27:23',  
      '17:27:24',  
      '17:27:25',  
      '17:27:27',  
      '17:27:28',  
      '17:27:29',  
      '17:27:31',  
      '17:27:32',  
      '17:27:34',  
      '17:27:35',  
      '17:27:36',  
      '17:27:37',
```

'17:27:38',
'17:27:40',
'17:27:46',
'17:27:48',
'17:27:49',
'17:27:50',
'17:27:52',
'17:27:53',
'17:27:54',
'17:27:55',
'17:27:57',
'17:27:58',
'17:27:59',
'17:28:01',
'17:28:02',
'17:28:03',
'17:28:04',
'17:28:05',
'17:28:07',
'17:28:09',
'17:28:10',
'17:28:11',
'17:28:13',
'17:28:14',
'17:28:16',
'17:28:17',
'17:28:18',
'17:28:19',
'17:28:21',
'17:28:22',
'17:28:23',
'17:28:24',
'17:28:25',
'17:28:27',
'17:28:28',
'17:28:32',
'17:28:33',
'17:28:35',
'17:28:36',
'17:28:37',
'17:28:38',
'17:28:39',
'17:28:41',
'17:28:42',
'17:28:43',
'17:28:44',
'17:28:45',

'17:28:47',
'17:28:49',
'17:28:51',
'17:28:52',
'17:28:54',
'17:28:55',
'17:28:56',
'17:28:57',
'17:28:58',
'17:29:01',
'17:29:03',
'17:29:04',
'17:29:05',
'17:29:06',
'17:29:08',
'17:29:09',
'17:29:10',
'17:29:11',
'17:29:12',
'17:29:14',
'17:29:15',
'17:29:16',
'17:29:17',
'17:29:18',
'17:29:20',
'17:29:21',
'17:29:22',
'17:29:23',
'17:29:24',
'17:29:26',
'17:29:27',
'17:29:28',
'17:29:29',
'17:29:31',
'17:29:32',
'17:29:33',
'17:29:34',
'17:29:36',
'17:29:37',
'17:29:38',
'17:29:39',
'17:29:41',
'17:29:42',
'17:29:43',
'17:29:44',
'17:29:45',
'17:29:47',

'17:29:49',
'17:29:50',
'17:29:52',
'17:29:53',
'17:29:54',
'17:29:56',
'17:29:57',
'17:30:04',
'17:30:11',
'17:30:12',
'17:30:14',
'17:30:15',
'17:30:17',
'17:30:18',
'17:30:19',
'17:30:21',
'17:30:23',
'17:30:24',
'17:30:25',
'17:30:26',
'17:30:27',
'17:30:29',
'17:30:30',
'17:30:31',
'17:30:32',
'17:30:33',
'17:30:35',
'17:30:36',
'17:30:37',
'17:30:39',
'17:30:40',
'17:30:41',
'17:30:43',
'17:30:44',
'17:30:45',
'17:30:46',
'17:30:47',
'17:30:49',
'17:30:51',
'17:30:52',
'17:30:56',
'17:30:58',
'17:30:59',
'17:31:00',
'17:31:01',
'17:31:02',
'17:31:04',

'17:31:05',
'17:31:06',
'17:31:07',
'17:31:08',
'17:31:10',
'17:31:11',
'17:31:12',
'17:31:15',
'17:31:16',
'17:31:17',
'17:31:18',
'17:31:20',
'17:31:21',
'17:31:22',
'17:31:23',
'17:31:25',
'17:31:26',
'17:31:27',
'17:31:28',
'17:31:30',
'17:31:32',
'17:31:33',
'17:31:34',
'17:31:35',
'17:31:36',
'17:31:38',
'17:31:39',
'17:31:41',
'17:31:42',
'17:31:43',
'17:31:44',
'17:31:46',
'17:31:47',
'17:31:49',
'17:31:50',
'17:31:51',
'17:31:52',
'17:31:54',
'17:31:55',
'17:31:56',
'17:31:57',
'17:31:59',
'17:32:01',
'17:32:02',
'17:32:03',
'17:32:11',
'17:32:12',

'17:32:14',
'17:32:21',
'17:32:22',
'17:32:29',
'17:32:31',
'17:32:32',
'17:32:39',
'17:32:40',
'17:32:41',
'17:32:42',
'17:32:43',
'17:32:45',
'17:32:46',
'17:32:47',
'17:32:48',
'17:32:49',
'17:32:51',
'17:32:52',
'17:32:54',
'17:32:55',
'17:32:56',
'17:32:57',
'17:32:58',
'17:32:59',
'17:33:01',
'17:33:02',
'17:33:03',
'17:33:05',
'17:33:06',
'17:33:07',
'17:33:08',
'17:33:10',
'17:33:11',
'17:33:12',
'17:33:13',
'17:33:14',
'17:33:16',
'17:33:17',
'17:33:18',
'17:33:20',
'17:33:21',
'17:33:22',
'17:33:24',
'17:33:25',
'17:33:26',
'17:33:27',
'17:33:28',

'17:33:29',
'17:33:31',
'17:33:32',
'17:33:33',
'17:33:35',
'17:33:36',
'17:33:37',
'17:33:38',
'17:33:39',
'17:33:41',
'17:33:42',
'17:33:44',
'17:33:45',
'17:33:46',
'17:33:47',
'17:33:49',
'17:33:50',
'17:33:51',
'17:33:52',
'17:33:55',
'17:33:56',
'17:33:57',
'17:33:58',
'17:34:00',
'17:34:01',
'17:34:02',
'17:34:03',
'17:34:04',
'17:34:06',
'17:34:07',
'17:34:08',
'17:34:09',
'17:34:10',
'17:34:12',
'17:34:13',
'17:34:14',
'17:34:15',
'17:34:16',
'17:34:18',
'17:34:19',
'17:34:20',
'17:34:21',
'17:34:28',
'17:34:29',
'17:34:33',
'17:34:34',
'17:34:36',

'17:34:39',
'17:34:47',
'17:41:12',
'17:42:58',
'17:43:00',
'17:43:02',
'17:43:04',
'17:43:05',
'17:43:07',
'17:43:09',
'17:43:10',
'17:43:12',
'17:43:13',
'17:43:15',
'17:43:17',
'17:43:18',
'17:43:20',
'17:43:21',
'17:43:23',
'17:43:25',
'17:43:26',
'17:43:28',
'17:43:29',
'17:43:31',
'17:43:33',
'17:43:34',
'17:43:36',
'17:43:38',
'17:43:39',
'17:43:41',
'17:43:42',
'17:43:44',
'17:43:46',
'17:43:47',
'17:43:49',
'17:43:50',
'17:43:52',
'17:43:54',
'17:43:55',
'17:43:57',
'17:43:58',
'17:44:00',
'17:44:02',
'17:44:04',
'17:44:06',
'17:44:07',
'17:44:09',

'17:44:10',
'17:44:12',
'17:44:15',
'17:44:16',
'17:44:21',
'17:44:22',
'17:44:24',
'17:44:34',
'17:44:36',
'17:44:38',
'17:44:39',
'17:44:41',
'17:44:42',
'17:44:44',
'17:44:46',
'17:44:47',
'17:44:49',
'17:44:50',
'17:44:52',
'17:44:53',
'17:44:55',
'17:44:57',
'17:44:58',
'17:45:00',
'17:45:02',
'17:45:03',
'17:45:05',
'17:45:07',
'17:45:08',
'17:45:10',
'17:45:12',
'17:45:14',
'17:45:15',
'17:45:17',
'17:45:19',
'17:45:20',
'17:45:22',
'17:45:23',
'17:45:25',
'17:45:27',
'17:45:28',
'17:45:30',
'17:45:31',
'17:45:33',
'17:45:35',
'17:45:36',
'17:45:38',

'17:45:39',
'17:45:41',
'17:45:43',
'17:45:44',
'17:45:46',
'17:45:47',
'17:45:49',
'17:45:51',
'17:45:52',
'17:45:54',
'17:45:55',
'17:45:57',
'17:45:59',
'17:46:00',
'17:46:02',
'17:46:03',
'17:46:05',
'17:46:07',
'17:46:09',
'17:46:11',
'17:46:12',
'17:46:14',
'17:46:16',
'17:46:17',
'17:46:19',
'17:46:21',
'17:46:23',
'17:46:25',
'17:46:26',
'17:46:28',
'17:46:29',
'17:46:31',
'17:46:33',
'17:46:34',
'17:46:36',
'17:46:37',
'17:46:39',
'17:46:41',
'17:46:42',
'17:46:44',
'17:46:45',
'17:46:47',
'17:46:48',
'17:46:56',
'17:46:57',
'17:46:59',
'17:47:00',

'17:47:02',
'17:47:04',
'17:47:05',
'17:47:07',
'17:47:09',
'17:47:11',
'17:47:12',
'17:47:14',
'17:47:16',
'17:47:23',
'17:47:24',
'17:47:26',
'17:47:27',
'17:47:29',
'17:47:31',
'17:47:32',
'17:47:34',
'17:47:36',
'17:47:39',
'17:47:40',
'17:47:42',
'17:47:43',
'17:47:45',
'17:47:46',
'17:47:48',
'17:47:50',
'17:47:51',
'17:47:53',
'17:47:55',
'17:47:57',
'17:47:58',
'17:48:00',
'17:48:02',
'17:48:03',
'17:48:06',
'17:48:08',
'17:48:09',
'17:48:12',
'17:48:16',
'17:48:18',
'17:48:19',
'17:48:21',
'17:48:23',
'17:48:24',
'17:48:26',
'17:48:27',
'17:48:29',

'17:48:31',
'17:48:39',
'17:48:40',
'17:48:42',
'17:48:43',
'17:48:45',
'17:48:50',
'17:48:52',
'17:48:53',
'17:48:55',
'17:48:57',
'17:48:59',
'17:49:01',
'17:49:03',
'17:49:05',
'17:49:07',
'17:49:08',
'17:49:10',
'17:49:13',
'17:49:14',
'17:49:16',
'17:49:18',
'17:49:20',
'17:49:21',
'17:49:23',
'17:49:24',
'17:49:26',
'17:49:28',
'17:49:29',
'17:49:31',
'17:49:33',
'17:49:34',
'17:49:36',
'17:49:38',
'17:49:39',
'17:49:41',
'17:49:42',
'17:49:44',
'17:49:46',
'17:49:47',
'17:49:49',
'17:49:50',
'17:49:53',
'17:49:54',
'17:50:01',
'17:50:03',
'17:50:04',

'17:50:06',
'17:50:08',
'17:50:09',
'17:50:11',
'17:50:12',
'17:50:15',
'17:50:17',
'17:50:19',
'17:50:20',
'17:50:22',
'17:50:25',
'17:50:27',
'17:50:28',
'17:50:31',
'17:50:32',
'17:50:34',
'17:50:36',
'17:50:43',
'17:50:44',
'17:50:46',
'17:50:48',
'17:50:49',
'17:50:51',
'17:50:53',
'17:50:54',
'17:50:56',
'17:50:58',
'17:50:59',
'17:51:02',
'17:51:03',
'17:51:05',
'17:51:06',
'17:51:08',
'17:51:10',
'17:51:11',
'17:51:13',
'17:51:14',
'17:51:16',
'17:51:18',
'17:51:20',
'17:51:21',
'17:51:23',
'17:51:24',
'17:51:26',
'17:51:30',
'17:51:32',
'17:51:34',

'17:51:36',
'17:51:38',
'17:51:39',
'17:51:41',
'17:51:42',
'17:51:44',
'17:51:46',
'17:51:47',
'17:51:49',
'17:51:51',
'17:51:52',
'17:51:54',
'17:51:55',
'17:51:57',
'17:51:59',
'17:52:00',
'17:52:02',
'17:52:03',
'17:52:05',
'17:52:07',
'17:52:08',
'17:53:24',
'17:53:27',
'17:53:29',
'17:53:31',
'17:53:32',
'17:53:34',
'17:53:35',
'17:53:37',
'17:53:39',
'17:53:40',
'17:53:42',
'17:53:44',
'17:53:46',
'17:53:48',
'17:53:49',
'17:53:51',
'17:53:52',
'17:53:54',
'17:53:56',
'17:53:57',
'17:53:59',
'17:54:00',
'17:54:02',
'17:54:04',
'17:54:05',
'17:54:07',

'17:54:08',
'17:54:10',
'17:54:11',
'17:54:13',
'17:54:15',
'17:54:16',
'17:54:18',
'17:54:20',
'17:54:22',
'17:54:24',
'17:54:25',
'17:54:27',
'17:54:28',
'17:54:35',
'17:54:37',
'17:54:39',
'17:54:41',
'17:54:42',
'17:54:44',
'17:54:45',
'17:54:47',
'17:54:49',
'17:54:50',
'17:54:52',
'17:54:53',
'17:54:55',
'17:54:57',
'17:54:58',
'17:55:00',
'17:55:01',
'17:55:03',
'17:55:05',
'17:55:06',
'17:55:11',
'17:55:12',
'17:55:14',
'17:55:16',
'17:55:17',
'17:55:24',
'17:55:26',
'17:55:28',
'17:55:29',
'17:55:31',
'17:55:32',
'17:55:34',
'17:55:36',
'17:55:38',

'17:55:39',
'17:55:47',
'17:55:48',
'17:55:50',
'17:55:51',
'17:55:53',
'17:55:54',
'17:55:56',
'17:55:58',
'17:55:59',
'17:56:01',
'17:56:02',
'17:56:04',
'17:56:06',
'17:56:07',
'17:56:09',
'17:56:11',
'17:56:12',
'17:56:14',
'17:56:15',
'17:56:17',
'17:56:19',
'17:56:20',
'17:56:22',
'17:56:24',
'17:56:25',
'17:56:27',
'17:56:28',
'17:56:30',
'17:56:32',
'17:56:33',
'17:56:35',
'17:56:36',
'17:56:38',
'17:56:40',
'17:56:42',
'17:56:43',
'17:56:45',
'17:56:46',
'17:56:48',
'17:56:50',
'17:56:52',
'17:56:53',
'17:56:56',
'17:56:57',
'17:56:59',
'17:57:00',

'17:57:02',
'17:57:04',
'17:57:05',
'17:57:07',
'17:57:08',
'17:57:10',
'17:57:12',
'17:57:13',
'17:57:15',
'17:57:17',
'17:57:18',
'17:57:20',
'17:57:21',
'17:57:23',
'17:57:25',
'17:57:26',
'17:57:28',
'17:57:30',
'17:57:31',
'17:57:33',
'17:57:35',
'17:57:36',
'17:57:38',
'17:57:40',
'17:57:42',
'17:57:43',
'17:57:45',
'17:57:46',
'17:57:48',
'17:57:50',
'17:57:52',
'17:57:53',
'17:57:55',
'17:57:56',
'17:58:03',
'17:58:05',
'17:58:07',
'17:58:09',
'17:58:11',
'17:58:12',
'17:58:14',
'17:58:15',
'17:58:17',
'17:58:19',
'17:58:20',
'17:58:22',
'17:58:23',

'17:58:25',
'17:58:27',
'17:58:28',
'17:58:30',
'17:58:32',
'17:58:33',
'17:58:35',
'17:58:37',
'17:58:39',
'17:58:40',
'17:58:42',
'17:58:43',
'17:58:45',
'17:58:47',
'17:58:48',
'17:58:50',
'17:58:53',
'17:58:55',
'17:58:56',
'17:58:58',
'17:59:00',
'17:59:02',
'17:59:03',
'17:59:05',
'17:59:09',
'17:59:11',
'17:59:13',
'17:59:14',
'17:59:16',
'17:59:17',
'17:59:19',
'17:59:21',
'17:59:22',
'17:59:24',
'17:59:26',
'17:59:27',
'17:59:29',
'17:59:31',
'17:59:39',
'17:59:41',
'17:59:44',
'17:59:46',
'17:59:48',
'17:59:50',
'17:59:52',
'17:59:53',
'17:59:55',

'17:59:57',
'18:00:05',
'18:00:07',
'18:00:10',
'18:00:12',
'18:00:13',
'18:00:15',
'18:00:17',
'18:00:18',
'18:00:20',
'18:00:21',
'18:00:23',
'18:00:25',
'18:00:26',
'18:00:28',
'18:00:29',
'18:00:31',
'18:00:33',
'18:00:34',
'18:00:36',
'18:00:37',
'18:00:39',
'18:00:41',
'18:00:42',
'18:00:44',
'18:00:46',
'18:00:48',
'18:00:49',
'18:00:51',
'18:00:52',
'18:00:54',
'18:00:56',
'18:00:57',
'18:00:59',
'18:01:01',
'18:01:02',
'18:01:04',
'18:01:21',
'18:01:23',
'18:01:25',
'18:01:27',
'18:01:29',
'18:01:38',
'18:01:40',
'18:01:42',
'18:01:44',
'18:01:46',

'18:01:47',
'18:01:49',
'18:01:51',
'18:01:53',
'18:01:54',
'18:01:56',
'18:01:58',
'18:02:00',
'18:02:02',
'18:02:04',
'18:02:05',
'18:02:07',
'18:02:08',
'18:02:10',
'18:02:12',
'18:02:13',
'18:02:15',
'18:02:17',
'18:02:19',
'18:02:21',
'18:02:23',
'18:02:25',
'18:02:27',
'18:02:29',
'18:02:30',
'18:02:32',
'18:02:33',
'18:02:35',
'18:02:37',
'18:02:44',
'18:02:46',
'18:02:47',
'18:02:50',
'18:02:52',
'18:02:54',
'18:02:55',
'18:03:03',
'18:03:05',
'18:03:07']

[]: