



TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY

Programmeerimise süvendatud algkursus ITI0140

2015



Teema

- Regulaaravaldised (ingl *regular expressions, regex*)



Dokumentatsioon

<http://docs.python.org/3/library/re.html>

<https://developers.google.com/edu/python/regular-expressions>



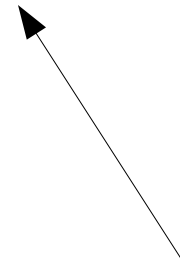
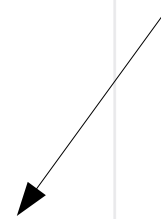
re

```
import re
```

```
match = re.search("abc", "aabca")  
print(match.group())
```

abc

Muster (mida otsitakse?)



Kust otsitakse? (Sõne)

Süntaks



		Tekst (kust otsitakse)	
Muster	Mustri näide	Leiab!	Ei leia!
-	"abc"	"aabc"	"ababac"
'.'	"a.bc"	"a2bc"	"abc"
'^'	"^abc"	"abcd"	"aabc"
'\$'	"abc\$"	"aaabc"	"abca"
'?'	"ab?c"	"aacc" "aabcc"	"aabbcc"
'*'	"ab*c"	"aaca" "aabca" "aabbcca"	"abab" "abbba"
'+'	"ab+c"	"aabca" "aabbcca"	"aaca" "abbba"

Süntaks



Muster	Mustri näide	Tekst (kust otsitakse)	
		Leiab!	Ei leia!
{m}	"ab{2}c"	"aabbcc"	"aacc", "aabcc", "aabbcc"
{m,n}	"ab{2,3}c"	"aabbcc", "aabbcc"	"aacc", "aabcc", "aabbcc"
\	"a\.bc"	"a.bc"	"a2bc"
[]	"[ab]cd"	"acd", "bcd"	"ccd"
	"[a-c]de"	"ade", "bde", "cde"	"dde"
	"[a\b]cd"	"acd", "bcd", "- cd"	"ccd"
(...)	"(ab)?cd"	"cde" "abcde"	"acde" "bcde"
	"(ab cd)?cd"	"acde", "abcde", "cdcde"	"ade"



Süntaks

`\w` - täht

`\s`- tühik, taandrida, reavaheetus jne

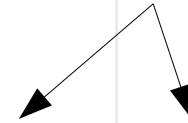
`\t \n \r` - taandrida, uus rida,
reavaheetus

`\d` - number



Grupid

Sulud tähistavad otsingugruppe



```
import re
```

```
match = re.search("xk(ab)?(cd)", "xkcde")  
print (match.group())  
print (match.group(0))  
print (match.group(1))  
print (match.group(2))
```

xkcd

xkcd

None

cd

Grupid



```
import re

match = re.search("(ab)?(cd)", "acd")
print (match.group())
print (match.group(0))
print (match.group(1))
print (match.group(2))
```

cd

cd

None

cd

Kui grupp ei leita, on selle
grupi väärtus None!



Grupid



(?:...) - vastet mittetagastav grupp

```
import re
```

```
match = re.search("(?:ab)?(cd)", "abcd")
```

```
print (match.group(0))
```

```
print (match.group(1))
```

```
match2 = re.search("(ab)?(cd)", "abcd")
```

```
print (match2.group(0))
```

```
print (match2.group(1))
```

abcd

cd

abcd

ab

compile()



```
regex_object = re.compile(".ab.")
result = regex_object.match("baba")
print(result.group())
result = regex_object.match("naba")
print(result.group())
result = regex_object.match("nab")
print(result.group())
```

baba

naba

Traceback (most recent call last):

File "", line 9, in <module>

print(result.group())

AttributeError: 'NoneType' object has no attribute 'group'

match() vs search()



match() otsib algusest

```
import re
a = re.match("c", "abcdef")
b = re.search("c", "abcdef")
print("Match" if a is not None else "No match")
print("Match" if b is not None else "No match")
```

No match

Match



findall()

```
import re
```

```
str = 'teretere vinge@gmail.com, koer jännes  
email@gmail.com ja nii edasi'
```

```
emails = re.findall("[\w\.-]+@[ \w\.-]+", str)  
for email in emails:  
    print(email)
```

```
vinge@gmail.com
```

```
email@gmail.com
```



Rise 4 Fun

<http://rise4fun.com/Rex>

nt "(ab|cd)?cd"



Regexi harjutamiseks

Regex Golf

<https://regex.alf.nu/>

**(Võite aine foorumisse
kirjutada oma skoorid.)**



Ülesanne

Ülesanne on nähtaval

<https://ained.ttu.ee>

<https://courses.cs.ttu.ee/pages/ITI0140>