

Regular expressions & Debug & other tools

Sugarkhuu Radnaa

Py4Econ in Ulaanbaatar

py4econ@gmail.com

Week 7: Learning objectives

Get to know:

- 1 Regular expressions
- 2 Debugging
- 3 Exception handling

Regular expressions

Search "pattern" from text and do an action Package "re":

- compile
- findall – find all matches
- match – match at the beginning of string
- search – first match anywhere in string
- split – split string into pieces at pattern points
- sub – replace match by user input
- group
- special characters

Regular expressions special characters

Regular Expressions Cheat Sheet by DaveChild

A quick reference guide for regular expressions (regex), including symbols, ranges, grouping, assertions and some sample patterns to get you started.

Anchors

^	Start of string, or start of line in multi-line pattern
\A	Start of string
\$	End of string, or end of line in multi-line pattern
\Z	End of string
\b	Word boundary
\B	Not word boundary
\<	Start of word
\>	End of word

Character Classes

\c	Control character
\s	White space
\S	Not white space
\d	Digit
\D	Not digit
\w	Word
\W	Not word
\x	Hexadecimal digit
\O	Octal digit

POSIX

[upper:]	Upper case letters
----------	--------------------

Quantifiers

*	0 or more	{3}	Exactly 3
+	1 or more	{3,}	3 or more
?	0 or 1	{3,5}	3, 4 or 5

Add a ? to a quantifier to make it ungreedy.

Escape Sequences

\	Escape following character
\Q	Begin literal sequence
\E	End literal sequence

"Escaping" is a way of treating characters which have a special meaning in regular expressions literally, rather than as special characters.

Common Metacharacters

^	[+	\$
{	*	(\
+)		?
<	>		

The escape character is usually \

Special Characters

\n	New line
----	----------

Groups and Ranges

.	Any character except new line (\n)
(a b)	a or b
(...)	Group
(?...)	Passive (non-capturing) group
[abc]	Range (a or b or c)
[^abc]	Not (a or b or c)
[a-q]	Lower case letter from a to q
[A-Q]	Upper case letter from A to Q
[0-7]	Digit from 0 to 7
\X	Group/subpattern number "x"

Ranges are inclusive.

Pattern Modifiers

g	Global match
i *	Case-insensitive
m *	Multiple lines
s *	Treat string as single line
x *	Allow comments and whitespace in pattern
e *	Evaluate replacement
U *	Ungreedy pattern
*	PCRE modifier

String Replacement

Debugging

In IDEs (VSCode):

- Actions at breakpoints: continue, step over, step into, step out
- Additional actions at breakpoints - expression, hit count, log message
- Watch
- Exceptions

Hand inputted breakpoint:

- `import pdb, pdb.set_trace()`

Exception handling

Exceptions occur from time to time in any sort of job. If you know what kind of errors potentially could occur and implement “remedy(ies)” on it beforehand

```
373
374     try:
375         # do main job
376     except:
377         # do another - Hereby, you are handling an exception!
378
```

Homework

- ① Task 1
 - ② Task 2
 - ③ Task 3
- Submit your result as a Github repository
 - Deadline: 1 week

Task 1: Regex

Өөрөө зохиох эсвэл бэлэн текст интернетээс олж regex-ийн дараах үйлдлүүдийг ашиглан текстүүд гаргаж авах жишээнүүд үзүүл (нэг бүр дээр нь бус хамтад нь ашиглаж болно. Гэхдээ доорх бүх тэмдгээс ядаж 1 удаа ашиглаарай)

- `[a-zA-Z0-9]`, `[a-z]`, `[A-Z]`, `[0-9]`
- `\d`, `\D`, `\w`, `\W`, `\s`
- `^`, `$`, `?`, `*`, `+`, `.`
- `{m,n}`, `{,n}`, `{m,}`, `{n}`
- Look behind, Look ahead, Negative look behind, Negative look ahead

Task 2: Exception

- ① Generate array of 1000 random integers (containing both negative and positive) in numpy
- ② Create a sub-array of the above with only the negative even numbers. To do that, do the following:
- ③ Loop one by one through this array ...
- ④ ...when negative odd, then raise “odd error” and continue to next loop,
- ⑤ ...when even but positive, raise “sign error” and continue to next loop ...
- ⑥ ...when “negative even”, then append your list.

Task 3: Debugging

- Debugging гэж юу вэ?
- Breakpoint – ийн ямар 3 төрөл (log message г.м) Vscode дээр байдаг вэ?
- Debug-ийн step into, step over, step out – ийн ялгааг тайлбарла

Thank you!