Semántica, Datos Conectados y Minería de Datos Textual

Expresiones regulares

Cristina Tîrnăucă

Dept. Matesco, Universidad de Cantabria

Facultad de Ciencias – Máster en Data Science

Regular Expressions: Disjunctions

- Letters inside square brackets

Pattern	Matches
[wW]oodchuck	woodchuck,Woodchuck
[1234567890]	Any digit

Regular Expressions: Disjunctions

- Letters inside square brackets

Pattern	Matches
[wW]oodchuck	woodchuck,Woodchuck
[1234567890]	Any digit

- Ranges:

Pattern	Matches
[A-Z]	An upper case letter
[a-z]	A lower case letter
[0-9]	A single digit

Regular Expressions: Disjunctions

- Letters inside square brackets

Pattern	Matches
[wW]oodchuck	woodchuck,Woodchuck
[1234567890]	Any digit

- Ranges:

Pattern	Matches
[A-Z]	An upper case letter
[a-z]	A lower case letter
[0-9]	A single digit

- Negations in disjunctions

(caret means negation only when first in [])

Pattern	Matches
[^A-Z]	Not an upper case letter
[^Ss]	Neither S nor s
[^e^]	Neither e nor ^
[a^b]	The pattern a caret b

Regular Expressions: Disjunctions

- Letters inside square brackets

Pattern	Matches
[wW]oodchuck	woodchuck,Woodchuck
[1234567890]	Any digit

- Ranges:

Pattern	Matches
[A-Z]	An upper case letter
[a-z]	A lower case letter
[0-9]	A single digit

- Negations in disjunctions

(caret means negation only when first in [])

Pattern	Matches
[^A-Z]	Not an upper case letter
[^Ss]	Neither S nor s
[^e^]	Neither e nor ^
[a^b]	The pattern a caret b

- The pipe | for disjunction

Pattern	Matches
yours|mine	yours or mine
a|b|c	a or b or c

Regular Expressions

Other Useful Symbols: ? * + . ^ $

Pattern	Matches
colou?r	Optional previous character
oo*h	0 or more previous characters
o+h	1 or more previous characters
beg.n	Any character except a newline
^[A-Z]	The beginning of a string
.$	The end of a string

More on using re in python:
`http://docs.python.org/3/library/re.html`