

# CSCA48 Winter 2016

WEEK 11 – Regular Expression

Bo(Kenny) Zhao

University of Toronto Scarborough

March 23, 2016

## Today's Plan

- Eat



Our tutorial average of term  
test #2 is ~63.3%

- Regular Expression
- Quiz #5 (group quiz, to be handed in before the end)
- Term Test #2

## Rules of Regex

Q: Is ... a valid regex?






1. The empty string is not a valid regex.
2. There are 4 regexes of length 1

'0', '1', '2', 'e'

3. There are 3 ways to form regexes of length greater than 1
  - (i) Append '\*' to a valid regex
  - (ii)  $'(r_1|r_2)'$
  - (iii)  $'(r_1.r_2)'$






## Examples

Q: Is '(0)' a valid regex?

1. The empty string is not a valid regex. 
2. There are 4 regexes of length 1  
'0', '1', '2', 'e' 
3. There are only 3 ways to form regexes of length greater than 1
  - (i) Append '\*' to a valid regex 
  - (ii) '(r<sub>1</sub>|r<sub>2</sub>)'      r<sub>1</sub> and r<sub>2</sub> are valid regexes 
  - (iii) '(r<sub>1</sub>.r<sub>2</sub>)' 






## Examples

Q: Is '1.2' a valid regex?

1. The empty string is not a valid regex. 
2. There are 4 regexes of length 1  
     '0', '1', '2', 'e' 
3. There are only 3 ways to form regexes of length greater than 1
  - (i) Append '\*' to a valid regex 
  - (ii)  $(r_1|r_2)$   $r_1$  and  $r_2$  are valid regexes 
  - (iii)  $(r_1.r_2)$   missing brackets

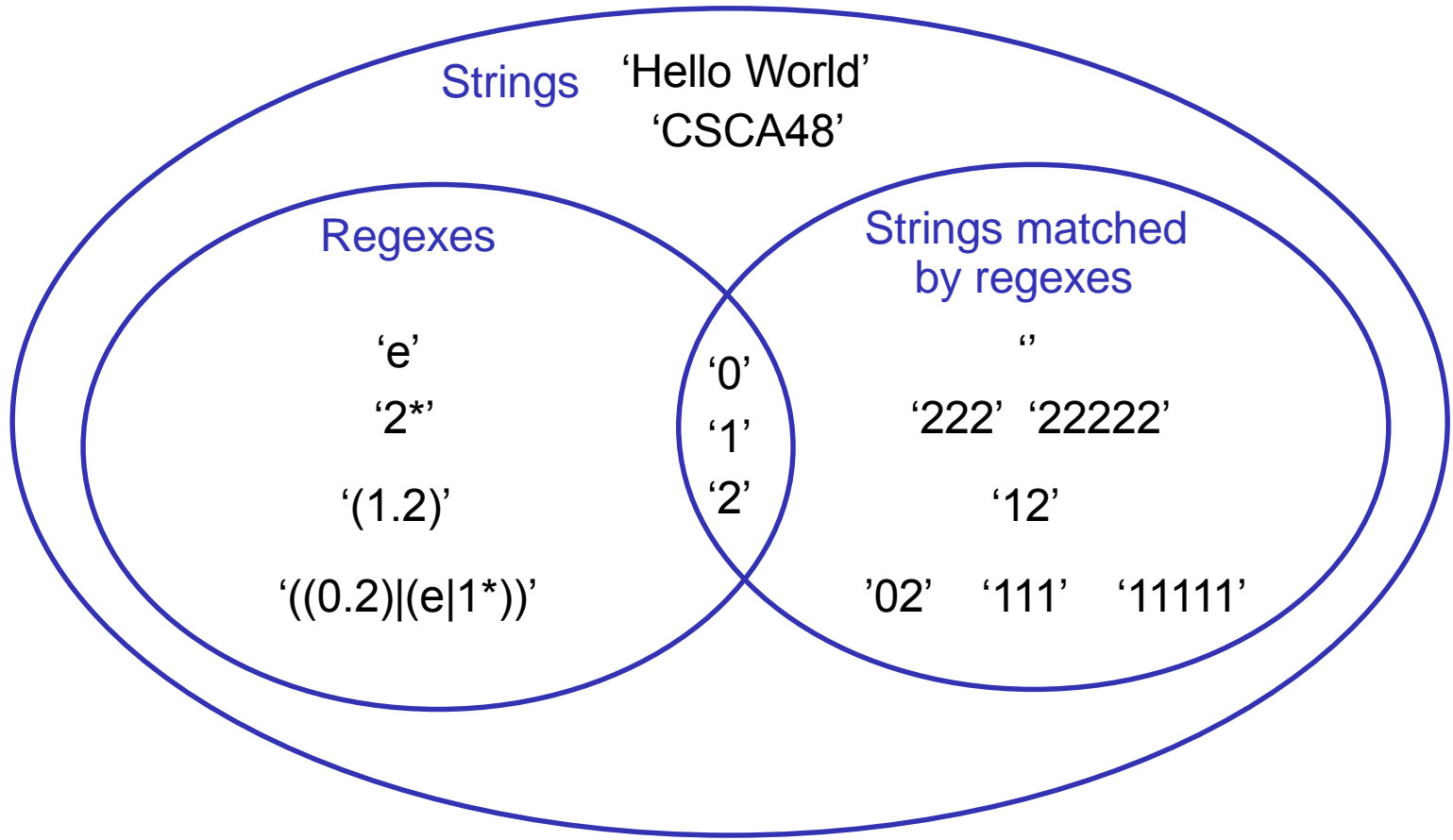
# Don't invent your own rules ...

A few rules you may come up with

- |  |                              |  |
|--|------------------------------|--|
|  | 1. $'*' + r$                 | $'*2'$ $'*1*'$ $'*(1.2)'$                                  |
|  | 2. $'(r_1 r_2 r_3)'$         | $'(0 1 2)'$ $'(0^* 1 0^*)'$                                |
|  | 3. $'(r_1.r_2.r_3 r_4.r_5)'$ | $'(0 1.2.e.1)'$ $'(0^*.1 0.2.e^*)'$                        |
|  | 4. $'(r)'$                   | $'(0)'$ $'(e^*)'$ $'((1.2))'$                              |
|  | 5. $' \quad r \quad '$       | $' \quad 0 \quad '$ $' \quad e^* \quad '$ $'(1.2) \quad '$ |

A valid regex with a few leading  
and/or trailing white spaces

## Relationship Between Strings and Regexes



# Regex Match

Q: Does ... match ... ?

1. '1' matches '1', '2' matches '2', '0' matches '0'
2. 'e' matches ''
3. '\*' means the regex it appends to may repeat 0 or many times

e.g. consider '1\*'

'1' repeats 0 times -> ''

'1' repeats 1 time -> '1'

'1' repeats 2 times -> '11'

'1' repeats 8 times -> '11111111'



## Regex Match

Q: Does ... match ... ?

4. '.' means you must choose both sides

e.g. '(1.2)'	->	'12'	'(0.0)'	->	'00'
'((1.2).3)'	->	'123'	'((e.e).e)'	->	''

5. '|' means you must choose one side

e.g. '(1 2)'	->	'1', '2'	'(0 0)'	->	'0'
'((1 2) 3)'	->	'1', '2', '3'	'((e e) e)'	->	''

## Regex Match

Q: Does ... match ... ?

4. '.' means you must choose both sides

e.g. '(1.2)'	->	'12'	'(0.0)'	->	'00'
'((1.2).3)'	->	'123'	'((e.e).e)'	->	''

5. '|' means you must choose one side

e.g. '(1 2)'	->	'1', '2'	'(0 0)'	->	'0'
'((1 2) 3)'	->	'1', '2', '3'	'((e e) e)'	->	''

## Quiz #5

Q:  $(0|1^*)$  matches ... ?

'0', '', '1', '11', '111', '1111111', etc.

Q:  $(2.2^*)$  matches ... ?

'2', '2222', '2222222222', etc.

Q:  $((0|1^*)|(2.2^*))$  matches ... ?

'0', '', '1', '11', '111', '1111111', '2', '2222', '2222222222', etc.

## Quiz #5

Q: Write a regular expression that matches all of the strings in set1(it may match other strings as well), but doesn't match any of the strings in set2.

$((0|1).(1.0)^*)$