Homework 3

Christopher Chapline

February 11, 2015

1 Exercise 3.1

1.1 Part a

$\Sigma * gc\Sigma *$

1.2 Part b

$(\Sigma * (g + \epsilon)(\Sigma - c))*$

1.3 Part c

$(g + c + (t((aa) + (ata) + (tt))*)*)$

1.4 Part d

$(c + a + t + (g(c(g + c + t + \epsilon) + \epsilon) + \epsilon))*$

2 Exercise 3.2

2.1 Part a

$([A - Z] + [a - z])([A - Z] + [a - z] + [0 - 9])*$

2.2 Part b

$[0 - 9] * .[0 - 9][0 - 9]*$

2.3 Part c

$\Sigma\Sigma * .\Sigma\Sigma * .((com) + (org) + (edu))$

2.4 Part d

$(\ \ + \backslash t + \backslash n)*$ where _

3 Exercise 3.3

First we need to find instances in which there are multiple newline characters. These are paragraph separations and should be marked. We mark them by assigning them a special character sequence (&& in this case):

replace(”$\backslash n \setminus n$”, ”&&”)

Now we replace singular newlines with a whitespace. This will condense simple line breaks into sentences:

replace(”$\backslash n$”, ” ”)

Lastly, we replace the previously substituted paragraph breaks with actual newline characters

replace(”&&”, ”$\backslash n \setminus n$”)

4 Exercise 3.4

4.1 Part a

$[A-Z][a-z]*' \ '([A-Z](.+[a-z]*)' \ ' + \epsilon)[A-Z][a-z]*(' \ 'Jr + \epsilon)$

4.2 Part b

$[A-Z][a-z]*, [A-Z][a-z]*((' \ '[A-Z](.+[a-z]*)) + \epsilon)$

4.3 Part c

$([0-9]([0-9]+\epsilon)' \ '[A-Z][a-z]*' \ '[0-9][0-9][0-9][0-9])+$
$([A-Z][a-z]*' \ '[0-9]([0-9]+\epsilon)' \ '[0-9][0-9][0-9][0-9])+$
$([[0-9][0-9][0-9][0-9]' \ '[A-Z][a-z]*' \ '[0-9]([0-9]+\epsilon))$

5 Exercise 3.5

5.1 Part a

replace(” or ”, ” && ”)
replace(”or ”, ”our ”)
replace(”&&”, ”or”)

replace(” er ”, ” && ”)
replace(”er ”, ”re ”)

replace(”&&”, ”er”)

5.2 Part b

For the ”or” → ”our” mapping, this conflicts with words like ”floor” and ”moor”. For the ”er” → ”re” mapping, flayer and computer conflict. These could all be handled by replacing them with custom strings prior to the primary replace calls. These custom strings would be changed back afterwards. This is similar to how the words ”or” and ”our” were handled.