Homework 3

Ka Man Chan

September 17, 2016

## Inititalize the raw data  
  
library(stringr)  
raw.data <-"555-1239Moe Szyslak(636) 555-0113Burns, C. Montgomery555-6542Rev. Timothy Lovejoy555 8904Ned Flanders636-555-3226Simpson, Homer5553642Dr. Julius Hibbert"   
name<- unlist(str\_extract\_all(raw.data,"[[:alpha:]., ]{2,}"))  
name

## [1] "Moe Szyslak" "Burns, C. Montgomery" "Rev. Timothy Lovejoy"  
## [4] "Ned Flanders" "Simpson, Homer" "Dr. Julius Hibbert"

#Question 1  
## Use the tools of this chapter to rearrange the vector so that all elements conform to the standard first\_name last\_name  
  
MyList<-c(",","Rev. ","Dr. ", "C. ")  
for (i in 1:4)  
 name<-str\_replace\_all(name, MyList[i],"")  
name

## [1] "Moe Szyslak" "Burns Montgomery" "Timothy Lovejoy"   
## [4] "Ned Flanders" "Simpson Homer" "Julius Hibbert"

#Question 2  
## Construct a logical vector indicating whether a character has a title (i.e. Rev. and Dr.).  
  
name<- unlist(str\_extract\_all(raw.data,"[[:alpha:]., ]{2,}"))  
Found\_Title<-str\_detect(name, "[[:alpha:]]{2,3}[.]")  
Found\_Title

## [1] FALSE FALSE TRUE FALSE FALSE TRUE

#Question 3  
## Construct a logical vector indicating whether a character has a second name  
  
name<- unlist(str\_extract\_all(raw.data,"[[:alpha:]., ]{2,}"))  
MyList<-c(",","Rev. ","Dr. ")  
for (i in 1:3)  
 name<-str\_replace\_all(name, MyList[i],"")  
Found\_Second\_Name<-str\_detect(name,"[[:alpha:]+.]{1,} [[:alpha:]+.]{1,} [[:alpha:]+.]{1,}")  
Found\_Second\_Name

## [1] FALSE TRUE FALSE FALSE FALSE FALSE

#Describe the types of strings that conform to the following regular expressions and construct an example that is matched by the regular expression  
   
#Question 4(1)  
## [0-9] Match any character in the set of 0-9  
## + The preceding item will be matchd one or more time  
## \\$ ends with $  
  
Ans4\_1 <- c("01234$")  
unlist(str\_extract\_all(Ans4\_1, "[0-9]+\\$"))

## [1] "01234$"

Ans4\_1

## [1] "01234$"

#Question 4(2)  
## \\b \\b Word boundary; Whole word only  
## Match any character in the set of a-z  
## min number of chararter is one; max number of charactor is four  
  
Ans4\_2 <- c("aazz")  
unlist(str\_extract\_all(Ans4\_2, "\\b[a-z]{1,4}\\b"))

## [1] "aazz"

Ans4\_2

## [1] "aazz"

#Question 4(3)  
## . Any character  
## \* The preceeding item (.) will be matched zero or more time  
## ? The preceeding item (.) is optional and will be matched at most one  
## \\.txt string .txt  
## $ match end of the string  
  
Ans4\_3 <- c("Question4\_3.txt")  
unlist(str\_extract\_all(Ans4\_3, ".\*?\\.txt$"))

## [1] "Question4\_3.txt"

Ans4\_3

## [1] "Question4\_3.txt"

#Question 4(4)  
## \\d{2} two Digits  
## with character /  
## \\d{2} two digits  
## with character /  
## \\d{4} four digits  
  
Ans4\_4 <- c("09/17/2016")  
unlist(str\_extract\_all(Ans4\_4, "\\d{2}/\\d{2}/\\d{4}"))

## [1] "09/17/2016"

Ans4\_4

## [1] "09/17/2016"

#Question 4(5)  
## < character <  
## ( Group starts  
## . Any character  
## + The preceeding item (.) will be matched one or more time  
## ? The preceeding item (.) is optional and will be matched at most one  
## ) group ends  
## > character >  
## . Any character  
## + The preceeding item (.) will be matched one or more time  
## ? The preceeding item (.) is optional and will be matched at most one  
## < character <  
## / character /  
## \\1 match the one in group (.+/?)  
## > character >  
  
Ans4\_5 <- c("<H1>Hello World</H1>")  
unlist(str\_extract\_all(Ans4\_5, "<(.+?)>.+?</\\1>"))

## [1] "<H1>Hello World</H1>"

Ans4\_5

## [1] "<H1>Hello World</H1>"

#Question 9  
## The following code hides a secret message. Crack it with R and regular expressions  
   
Original <- "clcopCow1zmstc0d87wnkig7OvdicpNuggvhryn92Gjuwczi8hqrfpRxs5Aj5dwpn0TanwoUwisdijLj8kpf03AT5Idr3coc0bt7yczjatOaootj55t3Nj3ne6c4Sfek.r1w1YwwojigOd6vrfUrbz2.2bkAnbhzgv4R9i05zEcrop.wAgnb.SqoU65fPa1otfb7wEm24k6t3sR9zqe5fy89n6Nd5t9kc4fE905gmc4Rgxo5nhDk!gr"  
  
## extract all string that match Upper case characters : A-Z, dot and exclaimation point   
secret\_msg <- unlist(str\_extract\_all(Original, "[[:upper:].! ]"))  
secret\_msg

## [1] "C" "O" "N" "G" "R" "A" "T" "U" "L" "A" "T" "I" "O" "N" "S" "." "Y"  
## [18] "O" "U" "." "A" "R" "E" "." "A" "." "S" "U" "P" "E" "R" "N" "E" "R"  
## [35] "D" "!"

## paste to concate the string  
## sep a character string "" to separate the terms  
## collapse an optional character string "" to separate the results  
secret\_msg <- paste(secret\_msg, sep="", collapse="")  
  
## replace character dot with a space " "  
secret\_msg <- str\_replace\_all(secret\_msg, "[\\.]", " ")  
secret\_msg

## [1] "CONGRATULATIONS YOU ARE A SUPERNERD!"