Digital Methods: Learning Journal

Mette Emily Kendon

Autumn 2020

1 02/11/2019

1.1 Thoughts/intentions:

My intention with this exercise was to see if I understand how to make use of the Regular Expressions on regex. At the hands-on-session we played around with the different ways you can combine and use the Regular Expressions, and I was at first a bit sceptical if I could figure out how to solve the exercise for today.

1.2 Action:

- The first exercise was to extract all the dates from a text.
- To find the text I went to Blackboard to find the slides from Thursdays hand-on-session. Here I clicked on the link http://bit.ly/regexexercise2 and copied all the text from this link and pasted it into regex.
- In the Regular Expression box i typed: (\d+).(\d+)..?(\d{4}|\b\d-\d-\d{4}\b) → this combination of Regular Expressions extracted all the dates from the text.
- The second exercise for today was to make our own stop-word lists for R and Voyant.
- To find the texts that I had to convert into two different stop lists I went to Blackboard to find the slides from Thursdays hand-on-session.
- Here, I first clicked on the link that led me to the Voyant list that I had to convert into a R stop-word list.
- I copied the Voyant list and pasted it into regex. Now it was time for me to use the Regular Expressions in order to convert the list into a list that can be used in R.
- In the regular expression box, I typed: (\n)|(\d)(^\b[a-zA-Z0-9ÆØÅæøå]\b) → This combination of Regular Expressions insured that I got a hold of all the words on the list including the foreign letters: ÆØÅ.
- To make the list applicable in R I inserted: ",$2$3" into the Substitution box. This made sure that the list was not horizontal, that each word was marked by quotations and that each word was separated by a comma.

- Afterwards I clicked on the link that led me to the R stop-word list that I had to convert into a Voyant stop-word list.
- I copied the R list and pasted it into regex.
- In the regular expression box, I typed: (", "| ") → this marked all the quotations and the commas.
- In the substitution box I typed: \n → this made sure that each word was on a new line without any quotations or punctuation.

1.3 Final thoughts:

Overall, I was a bit blown away by the many ways you can use Regular Expressions to search for information in large amounts of text, and it took me some time to figure out how to use them in this exercise. Eventually, I got the grip of how to use the different Regular Expressions, and the exercise turned out to be challenging in a fun way because it allowed me to play around with the different combinations. Finally, I got a better understanding of how a program like regex can be useful when working with large amounts of data because regex is a good tool for making these stop-word lists in a manageable way.