Teoriorienteret metode

W36 Regular expression – Rasmus Bøje Henriksen

1. **What regular expressions do you use to extract all the dates in this blurb:**[**http://bit.ly/regexexercise2**](http://bit.ly/regexexercise2)**and to put them into the following format YYYY-MM-DD?**
   * (\d{1,2}).(\d{1,2}).\s?(\d{2,4})

Link to Regex: <https://regex101.com/r/htOqQA/1>

1. **Write a regular expression to convert the stopwordlist (list of most frequent Danish words) from Voyant in**[**http://bit.ly/regexexercise3**](http://bit.ly/regexexercise3)**into a neat stopword list for R (which comprises "words" separated by commas, such as**[**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4)**). Then take the stopwordlist from R**[**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4)**and convert it into a Voyant list (words on separate line without interpunction)**

Form Voyant to stoplist for R: <https://regex101.com/r/11DJ7Y/1>

Form Stoplist for R to stoplist for Voyant: <https://regex101.com/r/fiKs9F/1>

1. **In 250 words, answer the following question: "What are the basic principles for using spreadsheets for good data organisation?",**

Spreadsheets are used, in a lot off different work environments and to at lot of different tasks. Spreadsheets can be used to data entry, storage, analysis, and visualization, in that regarding spreadsheets are a multipurpose tool. But as the text “Data Organization in Spreadsheets” points out it is important to try to separate primary data and the analyses/visualizations in different fields or spreadsheets. This is important so the data and calculations keep on, being transparent and so you easily can find the raw data and analysis. There is also other important principles is be consistent, don’t change underway in the spreadsheet. When you have given a type of category a specific name, sick to it. Like it’s important to be consistent, it is also important to be logical, when you are naming the fields, category and so on. Don’t name it, so it’s hard to find or not are transparent for others. This also goes for dates - go there with the “standard” for the field. Be carefully not to use extra space or special characters - this have often another meaning. Be logical when it comes to layout - this will help in the analysis. Maybe the most important principle is save the data, this can be done locale, but make backups. So if there data get corruptive you can go back. That’s why it’s also important not to do calculations in the raw data. It can also be useful to save data in a plain text file. This may not be pretty, but that way you can share the data none regarding the software.Øverst på formularen