INFO215 Seminar 5
Web Scraping with Scrapy

Today's plan:

Today we will work with the Scrapy python library that we can use to develop a more robust and scalable web scraper.

Task 1. Let us prepare for the next obligatory assignment 3. First look at the slides for lecture 5 and follow the steps there of installing/setting up Scrapy and then testing it with the 4 or 5 example codes (e.g article1.py) mentioned on the slides. If you can do this succesfully then it should be very clear what you need do to for the oblig and we also dont have to write a lot of additional code.

The 5 example codes that you should test your Scrapy installation with can be found on mitt.uib.no. When pip installing Scrapy you might get an error with a certain library, if the error tells you to install Microsoft C++ Package, then you should do so.

Make sure you understand what the code is actually doing. Remember that when you are running some of these example codes the code will scrape websites indefinitely which means that you must manually stop the program/terminal after a short while. Personally I close the process in my terminal by typing «ctrl + c». (sometimes I have to do it several times before it works).

Task 2. Try to solve the first part of obligatory assignment 3 which you can find on the end of the lecture 5 slides.

Task 3. If you have any remaining time, remember that we have quizzes on the INFO215 mitt.uib.no page that you can test yourself with to see if you have understood the topics of the week.