**Hands on exercise**

In this exercise you are asked to plan and implement a basic web crawler engine for news articles and implement a data extraction engine on the data.

The crawler engine is responsible for fetching and locally storing content from the internet.

The data extraction engine is responsible for extracting features from the content stored

1. Implement a web news scraper collecting titles from editions of online newspaper [www.dailymail.co.uk](http://www.dailymail.co.uk). The archive of this newspaper is in <https://www.dailymail.co.uk/home/sitemaparchive/index.html>.  
   for instance in October 21st 2018 first 5 titles were:

* Hayden Panettiere and her new beau Brian Hickerson hold hands as they hit theatre in Florida... months after split with Wladimir Klitschko
* Northern Ireland veteran `refusing medical treatment to have his day in court´
* 'Keep your mouth shut and your eyes open': Champion swimmer Emily Seebohm takes a swipe at ex-boyfriend Mitch Larkin after the Olympian denied cheating on her
* Vanessa Hudgens buys herself a bouquet of flowers at the farmers market in Los Angeles
* Teenager who beat his ex-girlfriend to death with a rock as a schoolyard dare to win a fried breakfast finally confesses as victim's mother blasts him for putting family through 'pain' of trial

1. Using the titles of the articles collected; generate a list of top 5 topics per month  
   Show results for each month of 2019.

**Additional remarks:**

* Use programing / scripting language of your choice.
* Explain the methodologies used in the data extraction engine and why these were chose.
* Work according to clean code methodologies (error handling, code structure, naming etc.)