

1. Date range is from 01-01-2007 to 21-08-2019
2. So called webfeatures were not included as they did not follow the general structure of articles
3. Scraping took a total of 9 hours split on two computers. In total this adds up to x amount of calls.

The files were concatenated using:

```
1
2 def append_files(files):
3     df = pd.concat([pd.read_csv(f, header = 0, sep = ';') for f in files],
4                     axis = 0)
5     df.to_csv('./data/{}_concat.csv'.format(files[0].split('.')[0]), index =
6               False) #Save to csv named as the first file with 'concat' appended
7
8 lst = ['dr_contents{}.csv', 'dr_links{}.csv', 'dr_log{}.csv']
9 for i in range(len(lst)):
10     append_files([lst[i].format(''), lst[i].format('2')])
```