**Pre-processing:**

* First we list only the directories from year > = 1900 of the GDL.
* Then, we iterate over all the folders in this list and open every file contained in each one of them. That is 12 xml files per year.
* After opening the files, we read them and then parse them using BeautifulSoup and its html parser.
* Then we retrieve all the names of the articles published during these years and store them in a list. We store the corresponding issue dates of these articles in another list.
* Once the above steps are executed, we create a data frame gathering these two lists. They show the name of the article and its issue date on the same row.
* We remove the rows with articles’ name « Untitled Article » for the moment.
* Using the pandas properties of a Series, we count the number of occurrences of the terms « Ecologie » and « Environnement » in the articles name columns of the dataframe. The output series allows us to plot histograms of the number of occurrences of these terms in regard of the year the article was published.