academic exchanges

P.Carrivain p.carrivain@gmail.com

November 27, 2018

The project of climatic vizualisation has the following tree :

```
globe/
 data/
 globe/
  data/
  js/
  index.html
 trade/
  data/
  js/
  index.html
 phase/
  data/
  js/
  index.html
 pubmed/
 scopus/
 pubmed_global_search.sh
 pubmed_search.sh
 merge.py
 scopus_global_search.sh
 scopus_search.sh
```

The bash script `pubmed_global_search.sh` works like :

```
./pubmed_global_search.sh -t <from> -T <to> -d <time-step> -w <keyword 1> -W <k
```

The bash script will request the pubmed database to count the number of papers with keyword 1 (keyword 2 is optionnal) in the title and the abstract for the period between "from" and "to" with "time-step" increment. The "restart" parameter is 1 if you would like to delete the previous request you asked for. The result of the request is a json file you analyse with the python script "merge.py". The json file for the title resquest (the abstract request) can be find in the folders :

```
pubmed/keyword/
pubmed/keyword/abstract/
```

The bash script `pubmed_search.sh` is a more detailed version of `pubmed_global_search.sh` and works like :

```
./pubmed_search.sh -t <from>
```

```
-T <to>
-d <time-step>
-w <keyword>
-r <restart>
-c <countries>
```

The bash script `pubmed_search.sh` uses the file `countries_ll.in` that can be find in the folder "data". The file `countries_ll.in` is an non-exhausting alphabetical order of the countries with the latitude, longitude and continent informations. The result of the request is a json file you analyse with the python script "merge.py". The json file for the title resquest can be find in the folder :

```
pubmed/keyword/pubmed_search_results_<from>/<country>/search_result_<country1>_
```

If you use the option :

```
-c countries
```

you will request the pairwize made of all the countries you have in the `countries_ll.in` file. If you specify a country you will request the pairwize made of this country and the one we find in `countries_ll.in` file. The pairwize are ordered such that the alphabetical order holds.

If you replace "pubmed" by "scopus" in the previous bash script we send request to the scopus database. However, the request number limitation is like twenty thousands per week. The scan of the scopus database per country is too much time.

The list of keywords can be find in the folders :

```
pubmed/title_abstract_keywords.in
scopus/title_abstract_keywords.in
```

If all the keywords you want to look for can be found in the file `pubmed/title_abstract_keywords.in` you can use the following bash command :

```
for w in $(more pubmed/title_abstract_keywords.in | cut -f1); do ./pubmed_glob
```

For a correlation search with keyword w_1 and keyword w_2 in the title/abstract you can use :

```
for w1 in $(more pubmed/main_keywords.out | cut -f1 -d ' ');
do
for w2 in $(more pubmed/main_keywords.out | cut -f1 -d ' ');
do
./pubmed_global_search.sh -t 1990 -T 2019 -d 1 -w $w1 -W $w2 -r 0;
done;
done
```

where the file `pubmed/main_keywords.out` contains all the very hot topics with more than the average of "IPCC publications" (in the pubmed and scopus databases) per year. If the previous condition is true for at least one year we consider building the correlations for all the years. The average of publications per year over the IPCC topics comes with the key `IPCC_mean_publications_per_year`

The python script "merge.py" is used to create a json file where we merged all the pubmed and scopus requests. We use it like :

```
python3 merge.py -t <from> -T <to> -d <time-step> -D <database>
```

The script writes the following json files :

```
globe/data/all.json
phase/data/title_abstract_pubmed.json
phase/data/title_abstract_scopus.json
```

You can load the `title_abstract_pubmed.json` and `title_abstract_scopus.json` files and get access to the keyword you want we the key ¡keyword¿. The number of keywords comes with the key `n_keywords` while the number of years our requests cover comes with the key `n_years`. The list of keywords (respectively the list of years) comes with the key `i_keywords` (respectively `years`) and the index key running from 0 to the number minus one. The bijection relation comes with the key `keywords_i`. The same bijection holds for the key `correlations_keywords`. The global view of request can be access with the key ¡title¿/¡abstract¿. The length of each key is accessible with `n_titles` and `n_abstracts`. You can then loop over the list and extract the year(s) and keyword(s) you are interested in. The global view of request about hot topics correlations can be access with the key `<double_title>/<double_abstract>`. You can remove from your hard-disk all the json files with a number of publications equals to zero with the bash script `clean_json_count0.sh`.

Globe vizualisation :

The globe vizualisation is based on the file `globe/data/all.json`. A version with only the data from pubmed can be found in `pubmed/big_json_pubmed.json` with the following tree :

1. keyword
2. pileups or links
3. year
4. country 1
5. country 2 (country 2 is always following country 1 according to alphabetical order)
6. the key "c" gives access to the number of publications

We use the python script `from_json_to_merge_json.py` to build the file.

Then with the key ¡pileups¿ you get access to a list (country,year,number of publications) related to the keyword. Then with the key ¡links¿ you get access to a list (country 1,country 2,year,number of publications) related to the keyword. The python script "merge.py" is used to extract data from FAO csv files. We scan the following database : Trade detailed matrix `http://www.fao.org/faostat/en/#data/TM`
The value of the export from A to B is different of the import of B from A. Crop residues `http://www.fao.org/faostat/en/#data/GA`
Burning crop residues `http://www.fao.org/faostat/en/#data/GB`
Burning savana `http://www.fao.org/faostat/en/#data/GH`
Burning biomass `http://www.fao.org/faostat/en/#data/GI`
Enteric fermentation `http://www.fao.org/faostat/en/#data/GE`
Manure left on pasture `http://www.fao.org/faostat/en/#data/GP`

Manure applied to soil http://www.fao.org/faostat/en/#data/GU
Synthetic fertilizers http://www.fao.org/faostat/en/#data/GY
Livestock manure http://www.fao.org/faostat/en/#data/EMN
Food supply http://www.fao.org/faostat/en/#data/CC
Food supply livestock and fish http://www.fao.org/faostat/en/#data/CL
Forestry production http://www.fao.org/faostat/en/#data/FO
Land use http://www.fao.org/faostat/en/#data/RL
Environment land use http://www.fao.org/faostat/en/#data/EL
Production livestock http://www.fao.org/faostat/en/#data/QA
Production livestock primary http://www.fao.org/faostat/en/#data/QL
Production livestock processed http://www.fao.org/faostat/en/#data/QP
Crops production http://www.fao.org/faostat/en/#data/QC
Crops processed http://www.fao.org/faostat/en/#data/QD
Trade crops livestock http://www.fao.org/faostat/en/#data/TP
Annual population http://www.fao.org/faostat/en/#data/OA
Temperature change http://www.fao.org/faostat/en/#data/ET
Pesticides use http://www.fao.org/faostat/en/#data/RP
Pesticides average use per area of cropland http://www.fao.org/faostat/en/#data/EP
Pesticides import/export http://www.fao.org/faostat/en/#data/RT
Fertilizers from http://www.fao.org/faostat/en/#data/RFN(RFB)
Fertilizers trade values http://www.fao.org/faostat/en/#data/RV
Environment fertilizers http://www.fao.org/faostat/en/#data/EF
Environment emissions intensities http://www.fao.org/faostat/en/#data/EI
Emissions intensities rice cultivation http://www.fao.org/faostat/en/#data/GR
Emissions intensities cultivated organic soils http://www.fao.org/faostat/en/#data/GV
Land emissions http://www.fao.org/faostat/en/#data/GL
Agricultural emissions http://www.fao.org/faostat/en/#data/GT

VIZUALISATIONS :

Once the "all.json" file has been produced we can have a exploratory look at the "phase" and "globe" vizualisation at the https://github.com/theatlasofdata/ecoflow.

The two vizualisations we propose here have to be consider as exploratory tools of climatic data we gathered. Before clicking the button of the vizualisation you would like to see, please consider reading the two small introduction paragraphs right below. We are actually working on a third vizualisation, a graph about the architecture of the hot topics.

PHASE PORTRAIT : We propose first a coarse-grained view of the academic exchanges on IPCC main keywords. Indeed, we look at the number of publications N per year with the requested keyword in the title. From there we can construct a publication velocity V between year Y-1 and year Y. Obviously, the total number of publications is growing fast and therefore we normalized our results by the total number of publications. We build then a phase portrait (a concept from the field of physics) as a tool to explore the dynamics of publications. The phase portrait does show the time as an implicit variable. Therefore, our tool draws a timeline evolution

of the normalized number of publications topic too. We clearly see an explosion of academic works about climate, greenhouse gases ... Eventually, we tackle the academic focus on the "Food and Agriculture Organization" (FAO) data. Indeed, the FAO provides great details on agriculture and associated CO2 emissions. Surprinsigly enough, the timeline evolution of the FAO academic concern is quite "chaotic". Nevertheless, a more detail vizualisation of the FAO data at the scale of country is far from without interest. Now that we request enough data to do some statistics we are working on potential publication dynamics between main keywords. We are currently building kind of a matrix of X-keywords correlations that will give us access to the differents levels of the IPCC main keywords architecture. Our first result is that the architecture of the IPCC main keywords in the framework of pubmed is limited to the two keywords correlations. Beyond, no academic exchanges have been extracted from the pubmed database.

Correlations between the hottest topics : We request the database to count the number of publications with two IPCC hottest topics in the title and in the abstract. The hottest topics are the topics with a number of publications per year greater than the average of the publications over the IPCC topics. The vizualisation we propose here is a a tool to explore the link between the hottest topics. You can display all the links an hottest topic make or just display the links between two of the main topics. The thickness of the link is directly related to the number of publications that is simply the synergy between the two topics. The center of the graph lies between the "mortality", "morbidity" and "climate" keywords.

GLOBE : After the coarse-grained view of the academic exchanges and its dynamic we focus on a more detail vizualisation of these data. We are now requesting the pubmed database about collaboration between researchers from differents countries. We end-up with a three dimensionnal graph where each link we extract from pubmed is shown as well as the strength of collaboration. The strength of collaboration is no more than a sphere travelling between the two collaborators with a speed going like the volume of common publications. You can easily navigate through the years and the main keywords to show the intersection between IPCC and Pubmed academic exchanges. The 3D graph about climate academic exchange does show a huge number of links. We can select the part of the distribution (the more important strength of collaborations) you want to study. You could also select only the links involving country or continent to "zoom in" the academic exchanges. For some of the IPCC main keywords this function is not useful. Indeed, subject like coral involves Australia as a main actor. As another example, the academic exchanges on the permafrost only hold for the north hemisphere countries. Eventually, we gathered some of the data from the "Food and Agriculture Organization" (FAO) that we included in the 3D globe vizualisation. From the "FAO raw data" we build some observables to link the agricultural yield of meat, rice, sugar ... to the greenhouse gaz emissions. We are still exploring the data to build more interesting observables about agriculture and climate at the scale of continent or at larger scale.