Open Climate Knowledge

About

Homepage: https://github.com/petermr/climate

Description:

Open Climate Knowledge (OCK) is an open research project for data mining Open Access (OA) papers and research outputs related to Climate Change — to build stats on research and OA rates, and for researcher to use in their work. OCK is intended for researchers inside and outside of academe.

> *Get involved! Run the software yourself or contribute to the research project. You can see open tasks and areas of expertise that need covering here. If you have questions or comments then raise an issue on GitHub or message on Twitter with the hashtag #OCK .*

The mission of OCK is to make research related to climate change 100% open. Currently research publishing related to climate change appear to be at rates of <30% open access.

Objectives (2019) create climate repository and knowledge graph.

Start: 9th Sept. 2019

Status: exploratory; open collaboration

Log

29 Oct - Presentation for Dienstagnachmittags-Kaffee

Slides: OCK for TIB Google Drive

Feedback from attendees and notes:

About 6/8 people attended

- If we want to show % of articles that are OA from total corpus then we need to be able to search as big a corpus as possible to know how much is closed. There was a suggestion that we might be able to access Fraunhofer's abstract collection from Scopus and search abstracts. They have some special collection.
- Want to collect citations that should Open Access and Open Science are beneficial. This seems obvious, but for all people this is not something they are necessarily aware of and can help in making case if it is at hand. Lambert Heller - tagging at #OA #OS here would be good https://www.zotero.org/groups/1838445/generation_r/items/collectionKey/G96ZZJ9N
- I need to collate a list of what ContentMine parts actually do, a feature list. Maybe this already exists, but essentially I reached the end of my ContentMine knowledge quite quickly and wasn't sure about what processes it was running.
- For people to be able to pick up from the very extensive list of 'Help Needed' requests in slides I need to be able to present the areas of help in a clearer way and how they relate to the development or functioning of the project. If this is done then it can be made a lot more relevant to bibliometric, developers, and climate scientists. For example: a software developer who want to contribute; someone want to build a dictionary; do some stemming in WikiData; offer a data source; relate to knowledge graph work
- It was mentioned that wanting to make a dictionary from CMIP6 makes it relevant to two interests, communities. I need to follow up Anette Ganske on this comment as lost track of the issues

Practical

- How can dictionaries be made? Do they need to be made as they are here https://github.com/petermr/climate/blob/master/dictionary/climate.xml as XML files, whether WikiData or Wikipedia.
- We need to document the stemming for dictionaries
- Link software downloads from ContentMine home page http://contentmine.org/
- How long does climate change dictionary take to do a 1000 article search and retrieval. https://github.com/petermr/climate/blob/master/dictionary/climate.xml

Actions

1. Setup vidconf with OSL+ colleagues and PMR https://terminplaner4.dfn.de/ and https://www.conf.dfn.de/
2. Collect base Open Science and OA citations
3. Build ContentMine feature list
4. Define and relate tasks where help needed on Open Climate Knowledge
5. Carry out speed test 1k article test on climate dictionary

Week 47, (18-22 Nov) 2019 - Launch Force11 Working Group

Need to prepare for this start:

- Writeup
- Setup for tasks: I would suggest Open Science MOOC bonus module as then that way we can make the instructions of how to use ContentMine
- Decide start date

- Setup on GitHub
- Set a first group vidconf call date