Riana Minocher: **/eco_minocher/EARS/scraped**

####SCRAPING ARTICLES####

The folder **/eco_minocher/EARS/fetched** contains a list of citations (one PDF = one citation).

The ***goal*** is to attend to every PDF, going through the following steps:

1) GENERATE CITATION KEY

2) MAKE FOLDER

3) SAVE BIBTEX

4) FILL IN YAML

5) CHECK BIBTEX vs. PDF

6) DRAG & RENAME PDF

7) DOUBLE-CHECK FOLDER

See detailed instructions below.

---

1) GENERATE CITATION KEY

Open one of the citations (= one of the PDFs) in:
/eco_minocher/EARS/fetched

Go to google scholar and search for the citation (copy & paste the title, if necessary also authors and year).

Click on the --> “ <-- marks below the search result. A small window called "Cite" will pop up.
In the bottom row of the Cite-window, click on "BibTeX" and you will be guided to a new page.

It will look something like:

```
@article{rose1955experimental,
  title={Experimental histories of culture},
  author={Rose, Edward and Felton, William},
  journal={American Sociological Review},
  volume={20},
  number={4},
  pages={383--392},
  year={1955},
  publisher={JSTOR}
}
```

—> this is the Bibtex citation for the PDF you are working on

---

---

2) MAKE FOLDER

Copy the ***citation key*** from the first row of the Bibtex citation. In above's example, the citation key is:
rose1955experimental
Generally speaking, the citation key is a string composed of:
[NameFirstAuthor]+[PublicationYear]+[Keyword]

Create a new folder in:
/eco_minocher/EARS/scraped

As folder name, paste the citation key. With above's example, you would end up with the following folder:
/eco_minocher/EARS/scraped/rose1955experimental

---

3) SAVE BIBTEX

Open a new file in Sublime Text (File >> New File).

Copy ALL content of the Bibtex page and paste the content into the newly created file.

Once more, copy the citation key.

Save the sublime text file with the extension .bib in the folder you created. As file name, paste the citation key.

In above's example, you would end up with the following file:
rose1955experimental.bib

---

4) FILL IN YAML

Make a copy of the yaml template in:
/eco_minocher/EARS
Move the copied yaml template to the citation folder, in above's example:
/eco_minocher/EARS/scraped/rose1955experimental

As the yaml's file name, paste the citation key.

In above's example, you would end up with the following file:
rose1955experimental.yaml

Fill in the yaml fields based on information in the ***PDF***. Also consult the variable list in:
/eco_minocher/EARS

NOTE: During scraping, ignore variables that are highlighted in green on the variable list!

IMPORTANT:
The following four variables REQUIRE an entry:
- doi
- title
- year
- journal

If information on one of those variables is not available in the PDF, the PDF was probably 'fetched' from a web page other than the journal's.
--> make an effort to retrieve the 'official' PDF from the journal's web page, e.g. through **scholar.google.com** or **doi.org**
--> download the 'official' PDF and put it into the appropriate publication folder; delete the old PDF and rename the new one

Consult the variable list for additional instructions!

---

5) CHECK BIBTEX vs. PDF

Cross-check all Bibtex information against the information in the PDF.

When doing that, make sure that the CONTENT is identical.
The information does NOT have to be in the exact same FORMAT!

For instance, the name of an author can be entered as:
L. Aplin or ***Aplin, Lucy***
--> same information, different format

The name of a journal can be abbreviated:
Proceedings of the National Academy of Sciences of the United States of America or ***PNAS***
--> same information, different format

If there are any differences in BIB vs. PDF regarding the CONTENT of the information:
--> apply changes to the BIB
go to the yaml and leave a *scraper_comments*: *"**updated .bib entry for [x]"*** (x = author, journal, issue, etc.)

EXAMPLES - LIST OF AUTHORS:
if the list of authors in the PDF doesn't match the list of authors in the BIB-file
--> consult the list of authors on the journal's web page;
- if applicable
--> apply changes to the BIB-file in accordance with the journal's web page
--> leave a respective comment in scraper_comments: *"**updated .bib entry for list of authors"***

EXAMPLES - PAGE NUMBERS:
if the page numbers in the BIB start with an 'e', then it refers to an 'e-journal'
--> leave the page numbers in the BIB as they are
if the page numbers in the BIB appear to be arbitrarily assigned:
--> apply changes to the BIB in accordance with the PDF

EXAMPLES - YEAR OF PUBLICATION:
if the year of publication in the PDF doesn't match the year of publication in the BIB-file
--> choose the *year when the article was actually published*
--> you may have to consult the journal's web page in order to determine the correct year
- if applicable
--> apply changes to the BIB-file - **only the year, NOT the citation key!**
--> leave a respective comment in scraper_comments: ***"updated .bib entry for year of publication"***
--> compile a list in 'notes.txt' of all cases where the citation key and the PDF feature different publishing dates

IMPORTANT:
Whatever changes you may apply to the BIB, ***NEVER change the citation key***!

Consult the variable list for additional instructions!

---

6) DRAG & RENAME PDF

Move the PDF to the new folder, in the example above:
/eco_minocher/EARS/scraped/rose1955experimental

One more time, copy the citation key.

Rename the PDF you've just moved by pasting the citation key.

---

7) DOUBLE-CHECK FOLDER

Continuing with the example above, you should now have the following folder:
/eco_minocher/EARS/scraped/rose1955experimental

Containing these files:
rose1955experimental.bib
rose1955experimental.pdf
rose1955experimental.yaml

If so, you can proceed with the next PDF in /**eco_minocher/EARS/fetched.**
