**Cite Network Project**

* **Retrieve the articles**

ISAAC GL OR ISAAC GLENN OR ISAAC GLI OR ISAAC GLL OR ISAAC GLYN OR ISAAC GLYNN OR ISAAC GLYNN L OR ISAAC GL OR ISAAC GLENN OR ISAAC GLI OR ISAAC GLL OR ISAAC GLYN OR ISAAC GLYNN OR ISAAC GLYNN L

Searched using the above string on WOS -> Cited References Search to retrieve articles that have cited Isaac’s work. The articles were downloaded in *Bibtex* file format.

* **Retrieve the abbreviations list**

The entries in the cited-references had their titles in abbreviated format which were replaced with the original titles from information provided [here](https://images-webofknowledge-com.proxyiub.uits.iu.edu/WOKRS527R13/help/WOS/A_abrvjt.html).

These webpages were scraped using *Beautiful Soup* to create a dictionary which could be used to abbreviated titles with original titles while processing files.

* **Processing files and Creating the CSV file**

The [code](https://github.com/pulkitmaloo/cite-network-project) to process the files works as follows

* *read\_files.py* reads articles from the files and works as a *generator* to yield articles one by one.
* *parse\_files.py* processes the articles in the following manner:
  + If the article’s author is Isaac then it is skipped as Isaac is citing his own work. There were 9 such articles.
  + In the cited-references if the articles have the keyword “Isaac”, ie, Isaac as the first author, then only these entries are kept and others are removed.
  + However, if no cited-references article has Isaac as the first author these entries are inserted as they are. These can be identified by little manual work. I tried to write a script for doing this manual work however, as we have unstructured data it didn’t do the job since it was hard to match articles.
  + The abbreviated titles in cited-references have been replaced with the original ones.
  + The articles are written to the Excel file using *Pandas Excel Writer.*
* **Manual Work**

When none of the articles in the cited-references field have Isaac as the first author, these entries were added as they were. These articles could be identified manually by the following process:

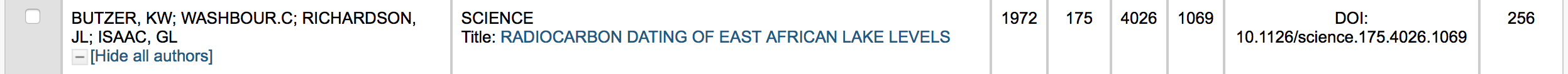
* + Let’s look at cited-references of article number 24 in Excel file

BUTZER KW, 1972, SCIENCE, V175, P1069, DOI 10.1126/science.175.4026.1069.

BUTZER KW, 1971, ENVIRONMENT ARCHEOLO.

Butzer K. W., 1968, DESERT RIVER NUBIA G.

* + Now let’s look at the articles we retrieved after searching on WOS in Step 2



* + Seeing the match, we identified the article by Isaac when he is not the first author. It’s the first one and we can replace its information and delete the rest of the articles in the cited-references field.