

FullTextRetrieval

Description

This script opens an Excel file of citation data, uses the NCBI E-utilities [ESearch API](#) to find PMIDs based on article titles, appends the PMIDs into link resolver URLs and writes those back into the spreadsheet. It uses Google Scholar or PubMed searches as fall-backs when the API does not find a unique PMID match.

[NCBI Disclaimer and Copyright Notice](#)

Information and best practices

This will work best with consistently formatted, named, and located spreadsheets; otherwise lines 15, 21, and 68 may need to be edited each time.

If the source spreadsheet includes a header row, remove `header=None` from line 15 and line 21.

The dataframe in line 21 could also be generated using `df = dforig.iloc[:, [3]]` where 3 is the index number of the title field. The code as written is used because it is easier to change from a human-readable standpoint.

The sleep function in line 62 adds a pause between each API request. By default, NCBI limits requests to 3 per second. That can be increased to 10 per second by obtaining an API key via an NCBI account. See E-utilities [Usage Guidelines and Requirements](#) for more information.

Prerequisites

The following Python packages need to be installed:

- [pandas](#)
- [requests](#)

To open the final URLs, install the "Open Multiple URLs" browser addon (available for both [Chrome](#) and [Firefox](#).) Select the "Do not load tabs until selected" box to minimize system strain. If there are a large number of citations in the file, open the URLs in batches of 20 or 25.

Attribution

Developed by Tiffany Tawzer and Christopher Beger.

ttawzer@luc.edu cbeger@luc.edu

Copyright 2023, Loyola University Chicago, distributed under the GNU General Public License version 3. The license text is available at <https://www.gnu.org/licenses/>

