

Homework 5 (aka Project)

Due: Monday, December 8, 2024 at 11:59 p.m.

Points: 100

Homework 4 had you produce a list of publication IDs from a keyword search on PubMed. The final project is to produce a list of the publication citations for that keyword.

Begin with your program 4. The program prints out a list of PubMed publication IDs. Use those to get the metadata. The URL for this is:

```
http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&retmode=xml&id=n
```

where n is a comma-separated list of PubMed publication IDs (or just a single PubMed publication ID). The web page you get back is an XML document giving details of the publications.

Your job is to print a bibliography from this record. Your entry for each journal should look like this:

A. Bester, R. Zelazny, and H. Ellison, "On the Role of Viruses in Future Epidemics," Journal of Future Results 3(4) pp. 29-35 (Mar. 2103). DOI 12.1119/2847595; PUBMED: 23456789.

Then print the abstract, if it is present in the record.

If there is no DOI, omit that part of the entry.

You will need to look at the XML records to get the fields. These are delimited by tags with attributes, each of which may have a value. For example, the element

```
<ELocationID EIdType="doi" ValidYN="Y">10.1016/j.vaccine.2015.04.071</ELocationID>
```

has a tag of `ELocationID`, attributes of `EIdType` (with value `doi`) and `ValidYN` (with a value of `Y`), and the field contains `10.1016/j.vaccine.2015.04.071`, which (as the `EIdType` value indicates) is a DOI. If its type were “pubmed”, the value would be the PubMed publication ID.

The fields of interest will have these tags:

- **Article** — contains the `Journal`, `ArticleTitle` (article title), `Pagination` (page numbers), `ElocationID`, which gives both the DOI and PII (if those exist), the `Abstract`, and the `AuthorList`.
- **Journal** — this consists of several elements, including `JournalIssue`, which contains the `Volume`, `Issue`, and `PubDate` (publication date), and `Title` (article title).
- **AuthorList** — this lists the authors, each author being in a field called `Author`. Subfields of interest are `LastName` and `Initial` (the initial of the first name)

If you want to see what the full structure is, just use the above URL to request a single item. This will also prove helpful in figuring out the attributes and tags you may need.

You can find methods for processing XML in the Python Library Reference at <https://docs.python.org/3.14/library/xml.etree.elementtree.html>