ISNI API notes 2017.11.06

Christine Fernsebner Eslao, HL ITS Metadata Management

APIs

An “Application Programming Interface” allows two software applications to talk to each other. In the case of both ISNI APIs, an application on one end of the interaction is a web service, and the other is a client -- we’ll start by using a web browser as our client, but later we’ll discuss how we might use other applications such as OpenRefine or Python scripts as clients.

A metaphor from one of the developers of OpenRefine:

Restaurant = API provider
Menu = API
Placing an order = executing an API call
Food items on menu = parameters of API call
Food delivered to your table = API response

One example of an pretty approachable API that’s in widespread use is Geonames’ web services: http://www.geonames.org/export/web-services.html

- Lookup by postal code (“02138”): http://api.geonames.org/postalCodeSearch?postalcode=02138&maxRows=10&username=demo
- Lookup by placename (“Cambridge”): http://api.geonames.org/searchJSON?q=%22Cambridge%22&maxRows=10&username=demo

Depending on how you “order” your “food”, you may get a different variety of dishes. Your software will need to be able to pick through these results and parse what you need from them. We’ll discuss these tools at the end.

Some characteristics of APIs that can vary:

- Response format(s)
- Need for registration, username, and/or authentication
- Frequency of calls allowed
- HTTP methods: GET, POST, PUT, DELETE
- Protocols, architecture, etc

ISNI SRU (“search/retrieve via URL”) API:

- XML responses
- No registration required
- GET

ISNI AtomPub (“Atom Publishing Protocol”) API:

- XML responses
- Membership & whitelisted IP address required
- POST

Demo

We’ll look at the search API in Chrome first:

- Use search API to make simple request by ISNI number: http://isni.oclc.nl/sru/?query=pica.isn+%3D+%220000000460275532%22&operation=searchRetrieve&recordSchema=isni-b
    - Base URL: http://isni.oclc.nl/sru/?
    - Parameters:
        - ISNI number: query=pica.isn+%3D+%220000000460275532%22
        - Operation type: &operation=searchRetrieve
        - Response schema: recordSchema=isni-b (“ISNI basic”)
- Use search API to make simple request by name keywords:
    - http://isni.oclc.nl/sru/?query=pica.nw+%3D+%22anne%20charlotte%20robertson%22&operation=searchRetrieve&recordSchema=isni-b
    - Name keywords parameter: ?query=pica.nw+%3D+%22anne%20charlotte%20robertson%22
- Use search API to make simple request by exact name
    - http://isni.oclc.nl/sru/?query=pica.na+%3D+%22robertson,%20anne%20charlotte%22&operation=searchRetrieve&recordSchema=isni-b
    - http://isni.oclc.nl/sru/?query=pica.na+%3D+%22robertson,%20anne%22&operation=searchRetrieve&recordSchema=isni-b
- Add parameters:
    - Maximum records to return: &maximumRecords=2
    - Year of birth: +and+pica.yob+>+"1940" (“after 1940”)
    - Check out other parameters in the form at http://isni.oclc.nl/sru/DB=1.2/

And then let’s look at one or more of these via HttpRequester. It’s best to try the AtomPub API using a tool like this (or script calls via Python, etc).

Resources

Documentation

ISNI SRU Search API: Guidelines and Examples
SRU standard: http://www.loc.gov/standards/sru/

ISNI AtomPub API Guidelines
ISNI AtomPub XML schema
AtomPub: https://tools.ietf.org/html/rfc5023

Tools

Web browsers

- Firefox plugin for making API calls and viewing responses: HttpRequester
- Chrome & Safari plugin for viewing XML: XV

Python

- Requests for making API calls
- Beautiful Soup for parsing XML
- iPython/Jupyter Notebook

If you don't feel comfortable with Python or don't need to collaborate:

- OpenRefine or RefinePro ease the creation of API calls based on spreadsheet data, and come with HTML parsing functions that are sufficient for basic XML parsing

Tutorials

Codecademy: APIs (tutorials for each specific API come with a "how to use APIs" lesson; coding languages vary)
OpenRefine: Fetching URLs from Web Services

Example script

https://github.com/eslao/isni_triage/blob/master/isni_matching.py

Task

Based on uncontrolled film directors' names from the Harvard Film Archive's internal database, create two spreadsheets:

- Names without matching ISNIs, so that catalogers can create ISNIs

- Names with one or more potentially matching ISNIs, so that catalogers can evaluate matches

Steps

1. Export director names, with film titles and other contextual data, from database into CSV
2. Create list of unique names
3. Query ISNI API and sort responses into “zero matches” and “one or more potential matches”
4. Write unmatched names into new CSV, with film titles and other contextual data
5. Write matched names into other new CSV, with film titles, other contextual data, total number of potential ISNI matches, and URIs for top five matches
6. Manual step: Upload to Google Sheets and share with catalogers

Results

Names without automated matches from sample no.1
Names with automated “matches” from sample no.1

Records in arbitrary sample of “archivist’s choice records from HFA database: 698
Number of unique name strings for directors in sample (excluding obvious non-names): 156

Number of strings with no ISNI matches via API: 35
Number of string with one or more potential ISNI matches via API: 121
Names with over 40 potential matches via API: “Robert Gardner” (44), “John Marshall with collaboration of Robert Gardner” (44), “John Carpenter” (46), “David Butler” (49), “James Boyd” (70), “John Marshall” (212), “John Smith” (544)