OpenHarvester

Description	Harvest data from open APIs for individuals, includes claiming interface and production of triples for VIVO
Type	ingest
Status	**PROOF OF CONCEPT**
Owner	Muhammad Javed
Language	Java, JavaFX
Team	Muhammad Javed
Location	https://github.com/mjaved495/OpenHarvester
License	

Goals

- Harvest data from open APIs for individuals, includes claiming interface and production of triples for VIVO. In future, same can be done for a paid subscription (such as WoS and Scopus).

Description

OpenHarvesters harvest publications metadata from different open source databases (CrossRef, PubMed and DBLP) and identify publications for a scholar. Name of the author can be mentioned differently, in the citation data of a publication of a source. For example, "Dean Blackmar Krafft", "Dean B. Krafft", "Dean Krafft", "DB Krafft", "D. Krafft" etc. These name variations make it harder to identify publications for a scholar using a named "search string". The difference between two distinct author names could be just the middle name initial e.g., "David F. Stern" and "David B. Stern". OpenHarvester algorithm learns from the claim publications as well as from the existing citation data of a claimed article. Currently, claimed publications can be stored in CSV, TXT, PDF and VIVO-Model format. No work is currently done in regards to "performance improvement" but focus was on "identification of correct publications".

Features

- User will download the zip folder on their system, unzip it and run it on their own system. Hope is to transform this application into a web application in the near future (looking for interested parties).

Documentation

Please see description on https://github.com/mjaved495/OpenHarvester

Notes

- None