MR04 Identify citations for research software in scientific articles

Background

For years, Graph Neural Networks have been growingly adopted for cases where data are not independent and identically distributed. Drug-protein prediction, social network clustering and scientific article/software recommendations are examples of applications where GNN has succeeded. However, it is not straightforward to model research articles and research software in the same graph, primarily since these last ones deal with complex heterogeneous metadata contents, formats, and source code containing both programming and natural language.

Goal

We aim to use Heterogeneous Graph Neural Networks to identify software in mathematical research articles, emphasizing software metadata. The approach must be then implemented in Julia using standard libraries.

Tasks

- Set up a Julia environment on your computer

- Build a pipeline to prepare swMATH and zbMATH (restricted to ArXiv sources) data and metadata

- Use standard libraries to inject the relevant data in a GNN model and train this model to identify software in scientific articles

Maxence Azzouz

Maxence.Azzouz-Thuderoz@
fiz-karlsruhe.de

Moritz Schubotz

Moritz.Schubotz@
fiz-karlsruhe.de

MR07 Mining information of swMATH software

Background

swMATH indexes software introduced by zbMATH articles. The metadata collected by zbMATH are essentially harvested by hand, with article as essential support to find information. However, intrinsic metadata of software are much unexplored and could considerably help with enriching the swMATH catalog of information.

Software Metadata Extraction Framework (SOMEF)

`docs passing` `python 3.9` `pypi package 0.9.5` `DOI 10.5281/zenodo.10848711` `launch binder` `repo status Active`

A command line interface for automatically extracting relevant metadata from code repositories (readme, configuration files, documentation, etc.).

Demo: See a demo running somef as a service, through the SOMEF-Vider tool.

Authors: Daniel Garijo, Allen Mao, Miguel Ángel García Delgado, Haripriya Dharmala, Vedant Diwanji, Jiaying Wang, Aidan Kelley, Jenifer Tabita Ciuciu-Kiss and Luca Angheluta.

Goal

We aim to use the somef library to extract intrinsic metadata of software repositories identified in swMATH

Tasks

- Install the packahe https://github.com/KnowledgeCaptureAndDiscovery/somef

- Build a pipeline in python to extract intrinsic metadata

- Compare the extrinsic metadata of swMATH and the intrinsic metadata harvested in software repositories

Maxence Azzouz

Maxence.Azzouz-Thuderoz@
fiz-karlsruhe.de

Moritz Schubotz

Moritz.Schubotz@
fiz-karlsruhe.de