Fondren Fellows Project

Any questions can be directed to fondrenfellows@rice.edu

Name
Arko Barman, Ph.D.

Project Title
Maintain and Analyze Rice University Research Repository

Proposal Description:

Rice University Research Repository (R3) is a database that contains the research outputs of Rice faculty and staff. Maintaining and analyzing the repository is a daunting task. This project will aim to develop software that can update the repository, analyze, and mine it for generating useful results, analyses, and visualizations.

Project Summary:

The proposed project aims at developing software for maintaining, updating, analyzing, and mining the R3 repository. The first task of the project would be to analyze the completeness of the repository with the aid of other open-source indexing resources for research output, e.g., Google Scholar. The project will then be aimed at trying to complete the database and creating a mechanism by which it can be updated and maintained using an easy-to-use software API. In parallel, the project will involve the development of software API for easy querying of the database for analysis, mining, and generating useful and meaningful visualizations that will not only be useful for the library personnel but also for the research community at Rice.

The project also encompasses research into the licensing and journal use policies for the agreements that Fondren has with academic publishers. Analysis of the different publisher policies and the barriers to R3 is another aspect of the project deliverables, which includes an enumeration of the key publishers, the number of Rice publications encompassed by each, and establishing a typology of the use restrictions for the publishers.

The fellows would work in conjunction and collaboration with the faculty mentor and possibly a team of students in the Data Science Capstone Program, who will focus on the statistical and machine learning analysis of the database. The fellows should be proficient in Python programming, and be familiar with databases, data mining, statistical analysis, and visualization. The project will aim to populate and complete the database with open-source information only, such as metadata and abstracts related to a publication.

The project will directly aid the Fondren Library in the maintenance and periodic updates to the R3 repository. In addition, it will create useful software for maintaining and accessing information on research outputs, such as publications and patents. The library personnel and the research community at Rice would also be able to query, analyze, and generate visualizations using the software.

The developed software would be helpful for the Office of Corporate and Foundation Relations in mining the database and query for the names of relevant faculty members as they serve as a liaison with corporate partners and foundations.

We propose to include Lisa Spiro as a mentor and liaison to help the fellow complete the project.

How many fellows are you requesting?

We are requesting 3 Fondren fellows.

Fellow 1: The fellow will work on analyzing the completeness of the R3 database and writing software to maintain and update it from time to time. They will use tools like Python and appropriate libraries for extracting information from Google Scholar or other similar aggregator websites to create a research profile for every Rice researcher on the R3 database that will include all their publications and patents. Additionally, they will be involved in the development of an efficient and effective database design for future use.

Fellow 2: The fellow would focus on analyzing the data in the repository for mining, analysis, and visualization. They will utilize the title, abstract, and other metadata from the publications and patents of every Rice researcher to create useful visualizations and create useful and tailored searches using data mining, information retrieval and AI tools.

Fellow 3: The third fellow would work on the analysis of the academic publisher agreements and the impact of those agreements on completing R3. They will distill the

information from publisher agreements and contracts to identify restrictions and limitations on the usage of full text for different purposes, e.g., developing analysis tools, data mining and text analysis and also for different use cases, such as the legality of making these tools available within Rice.

Key Tasks:

1. Analyzing the completeness of R3
2. Maintaining and updating R3 with new research output
3. Analyzing, mining, and generating useful visualizations
4. Writing software that can be used to maintain the repository in the future
5. Writing software that can aid in querying, analysis, mining, and visualization of the information in the database
6. Enumerating the academic publisher agreements and their impact on R3.

Qualifications for applicants:

Fellow 1: Python, designing and querying databases, statistical analysis
Fellow 2: Python, querying databases, visualization, statistical analysis, data mining
Fellow 3: Interest in library science/ academic publishing. Publishing contracts.

What would students learn through their participation in this project?

Students would learn to design, create, and maintain databases. In addition, they would be exposed to handling a real-world database with all its challenges, and how to distill data to perform meaningful analyses and visualizations. Further, the students would learn the fundamental principles of collaborative software development. Students would also learn more about academic publishing licensing and use agreements as well as software in the space of mining published work.