**CrossRef Event Data**

Team members: Darpan Shah, Salsabil Bakth, Mohammad Tahmid, Tabish Shaikh, Rihat Rahman

**Project Overview**

The scholarly reward system in higher education focuses on three main areas: research, teaching, and service. Throughout the years, the weights assigned to the three areas have changed and institutions put a greater emphasis on research. Researchers measure the achievements of their work by developing various indexes to categorize the output of scholars across domains. They use indexes such as the h-index or the JIF (journal impact factor) to measure their work.

An area of work that researchers use to measure their engagement on the web is Altmetrics. Altmetrics examines the amount of activity a scholarly document receives on various online platforms using APIs. API’s such as Crossref or Crossref Event Data search the web across 13 different platforms like Twitter, Facebook, Wikipedia, Reddit, and etc. looking for interactions with scientific papers.

Sources such as Almetric.com and PlumX have been used by scholars, however these sources do not use the full potential of CrossRef Event Data due to the complex data that is presented. CrossRef is the organization that assigns DOIs for research publications. DOIs are a unique ID for scholarly documents and they are tracked by the Crossref Event Data API. The data consists of millions of records which are documented in open source, public APIs. Through this project, scholars can access a variety of data through a free, open source web application.

**Project Purpose, Scope, and Objective**

The purpose of this project is to extend an existing open source web application that will allow users to search scientific publications and view the number of citations and events. Using the results, users will be able to determine the engagement rate of the publication on various online platforms.

The objective of this project is to provide an open source tool for researchers to see the impact of their research throughout multiple social media platforms. Users will be able to look up publications using a DOI, article name, journal, keywords, or author. In addition to that, users will have the option to explore publications by a university, country, and other factors. To search multiple publications at once, users will be able to upload a list of titles using CSV files or text files. Search results will be downloadable and they will include visualizations to help apprehend the data.

The scope of this project is to implement both the front-end and back-end of the web application. Front-end will include a user-friendly search interface to allow search using plain text or files, while search results will comprise tables and other data visualization tools. The back-end of this application will utilize MySQL databases that will store event and citation information, fetched from multiple APIs.

**Team Organization**

**Darpan Shah**

* Team Lead: Responsible for the delegation of work within the team. He is the primary contact between the team, GTA, client, and instructor.
* Database/Back-end Lead: Responsible for the design and development of the database and back-end of the web application. Provides guidance to other team members when handling the back-end.

**Mohammad Tahmid**

* Presentation Lead: Responsible for ensuring presentation materials meet the criteria. Reviews presentation materials for appropriate format and completion.

**Rihat Rahman**

* Python Lead: Responsible for the design and development of python scripts. Provides guidance to other team members when handling python scripts.

**Salsabil Bakth**

* Front-end Lead: Responsible for the design and development of the user interface and front-end. Provides guidance to other team members when handling the front-end.

**Tabish Shaikh**

* Documentation Lead: Ensures that all sections of the document to be submitted are completed, reviewed for grammatical errors, and are formatted appropriately.
* QA Lead: Responsible for ensuring that the testing of features and components of the web application meet the requirements and expectations of the client.

**Project Resolution Policies**

The team will be using a strike system to handle all issues that occur during this project's lifetime. Receiving a strike from the team lead will result in a team meeting to discuss the issue and have the team lead issue a resolution. The team lead will have the final verdict on the matter. If a member receives three strikes by the team lead then the GTA will be informed of the problem and a meeting will be held with all members present. If a resolution is not achieved then the professor will be informed to come to a final decision.

Team members are expected to complete their piece of code by the deadline set beforehand. It is the responsibility of the team member that has difficulty with the code to inform the team beforehand and come to a resolution. Failure to complete the code without any notice will result in a strike.

Team members are expected to communicate and reply to messages in a timely manner. If the team lead decides that a team member has not communicated in a timely manner, then the team member will receive a strike. If a team member has any issues, they are to inform the team and discuss a solution.

Team members are expected to attend all team, GTA, and client meetings throughout the project’s lifecycle. If a team member fails to inform the team lead that they are going to miss a meeting beforehand with a legitimate reason then they will receive a strike.

The project is an extension of a previous semester’s project that has a set technology stack. If there are issues with the technology choices then the team will meet to discuss the issue and have the team lead issue a verdict. If the team lead is unable to come to a decision then the GTA and client will be informed and issue a final verdict.

**Project Plan (Iterations, Project Schedule)**

The team meets weekly with the client, Dr. Bowman, on Fridays at 3:00pm. The meeting time with Saeid, the assigned GTA, will be on Thursdays at 1:00pm. The team has decided to meet at least three times a week internally over Discord. The times are 1:00pm on Mondays, Wednesdays, and Saturdays. The team will also meet on Fridays outside the client meeting time if needed.

The project development will be done in three phases separated equally throughout the semester. Each phase will include a set of features that the client has indicated as goals for the expansion of the existing project.

The Saturday meetings will include a recap of the work done by each team member and what tasks they will take on in the next week which will include improving the task done throughout the current week or starting a new task.

**Tentative Project Schedule**

* **Requirements Gathering (02/04/2021)**
  + Document that indicates the features and behaviors that are expected from our software
* **Phase I (02/09/2021):**
  + Gather event data respective to DOIs from Dr. Bowman.
  + Upload of CSV Files containing list of DOIs to see if they exist in the database
  + Download of data from search as zip file containing JSON or CSV
  + Create data visualizations (tables and graphs).
* **Design Specification (03/02/2021)**
  + Document that contains information and diagrams about the hardware and software architecture of our project
* **Phase II (03/09/2021):**
  + Add SciELO Data (02/20/2021)
    - Clean JSON files for sharing
    - Relational database map
  + Add OpenCitations Data (03/02/2021)
    - Clean JSON files for sharing
    - Relational database map
* **Phase III (03/30/2021):**
  + Adding Security to upload feature
  + Validating DOIs
  + Filter search by university, country, or domain
  + CRON jobs
* **Final Presentation (04/20/2021)**
  + Presenting the final product of our work from the semester.

**Configuration Management Plan**

GitHub will be used as the version-control system for this project. The team will manage changes to code by allowing each member to create their own branch off the master branch. Each member will then add additional functions to their separate branches. The member will then test the implementation of the newly added function, and then create a merge request. In the case of a code conflict, the team will work together to review the code and determine potential solutions.

In addition, the team will also use Discord to manage the project. Discord is primarily being used for communication, but each member will also use it to inform the rest of the team of updates they push to the master branch on GitHub. The team will also use it to store meeting notes after any meeting with the client or GTA, as well as after any internal team meetings.

**Technologies**

Front-end:

* HTML, CSS, JavaScript (Front-End)
* D3.js (Visualization Library)

Back-end:

* Python (Back-End)
* Flask (Framework)
* MySQL (Database)

API:

* Crossref
* Crossref Event Data
* OpenCitations
* Postman for API Testing

Technology stack of this project was determined by the client, based on the first version of the application.