CIVIC DIGITAL FELLOWSHIP

Project Metadata Tool

Shalin Brahmbhatt

Supervised by Dr. Heather Madray, Dr. David Beede, James Kriebel

Center for Enterprise Dissemination

United States Census Bureau

Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau.

Introduction

- **Center for Enterprise Dissemination**
    - Assures that the Census Bureau can effectively disseminate the maximum amount of high-quality data about the Nation's people and economy
    - Protect the confidentiality of respondents and the information they provide
    - Facilitates access to restricted-use data
- **Project Metadata Tool**
    - Seeks to match external FSRDC (Federal Statistical Research Data Center) research projects with publications
    - Began as a CDF project in 2021 and continuing into 2022, potentially 2023

Motivations

Make Projects Publicly Visible

- Publishing research project metadata
- Centralize location to view FSRDC projects and outputs
- Surfacing publications from the research projects (not the restricted data!)

United States® Census Bureau

Motivations

Highlight Census Research

- Two types of research
  - External and Internal
- External research:
  - Approval from Census Bureau
  - Restricted data usage
  - Identifying data presented in the research

United States® Census Bureau

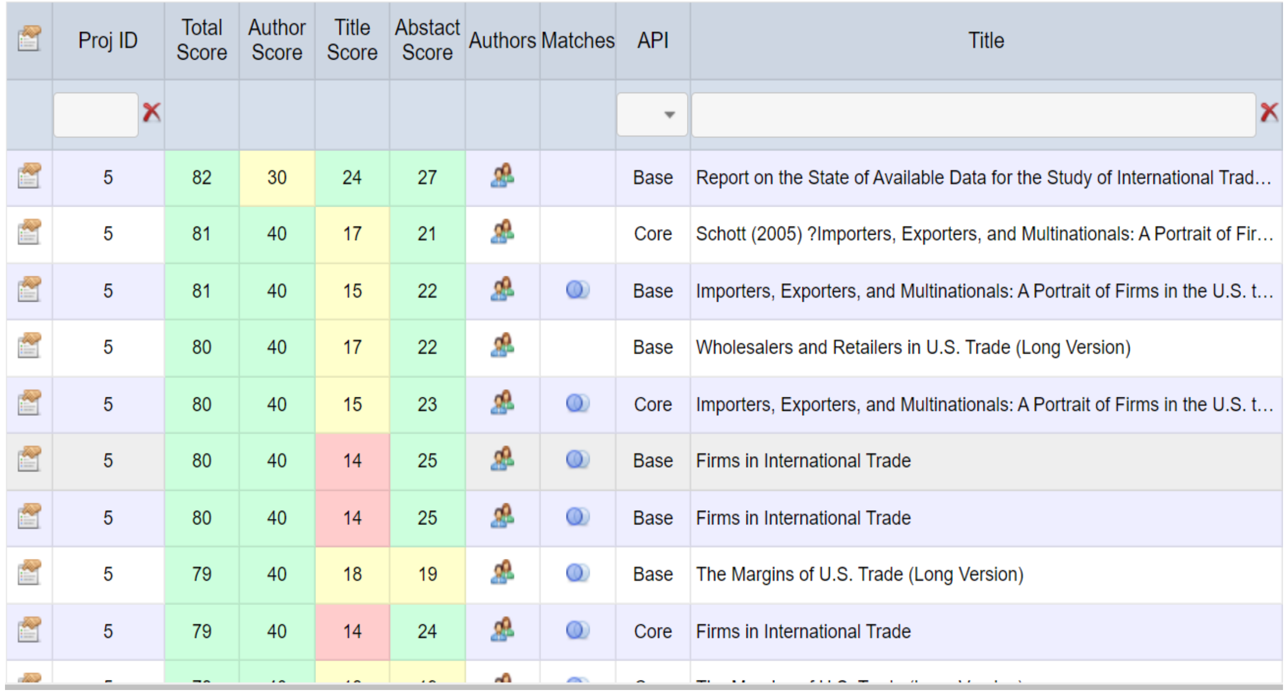
Truth Deck

- Prior to this project, the data had to be input manually
- Compare the abstracts and validities of research projects and publications by hand
- Approve or deny the pair of papers
- Hundreds of research projects to select from
- Thousands of publications from public sources per research project

United States® Census Bureau

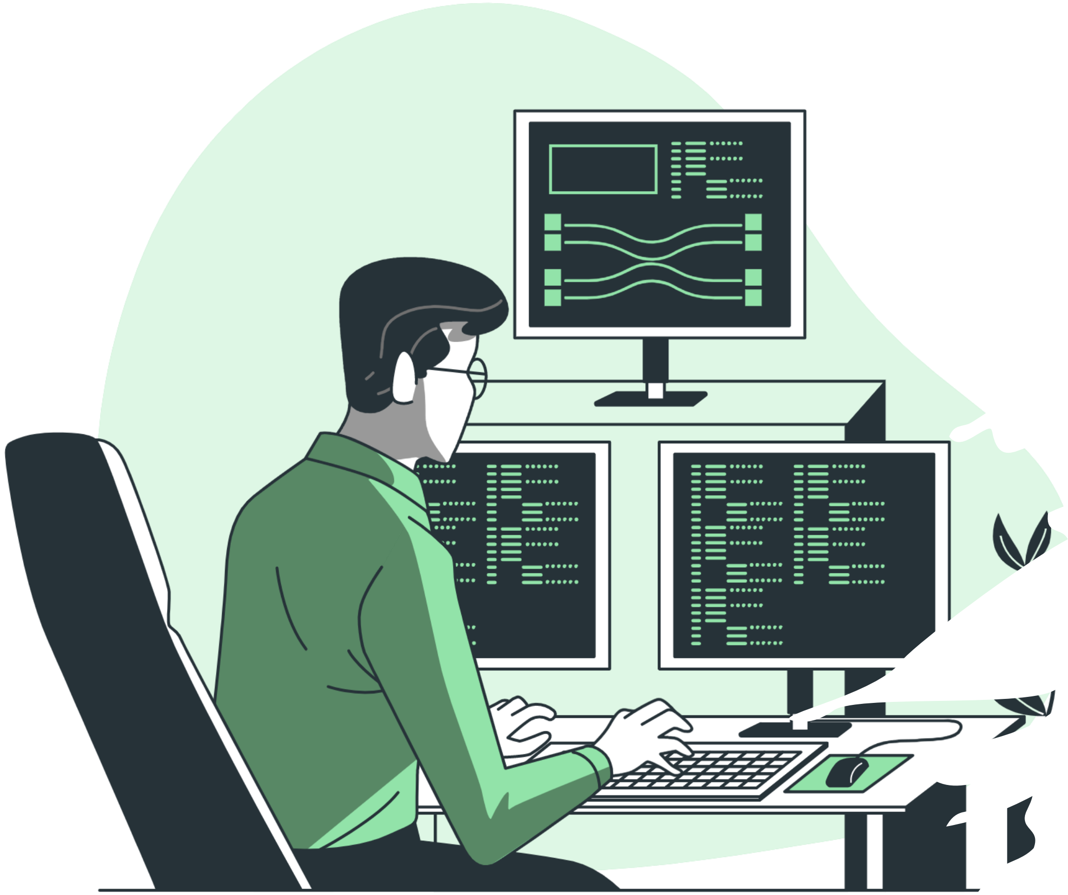
Scoring Mechanism

- How can we quantify the similarity between a research project and an associated abstract?
- Create a score to assign each publication, human-readable format
  - Start with the authors, title, and abstract
  - Work through different methods of analysis
  - Utilize the TruthDeck



📝	Proj ID	Total Score	Author Score	Title Score	Abstact Score	Authors	Matches	API	Title
📝	5	82	30	24	27	👥		Base	Report on the State of Available Data for the Study of International Trad…
📝	5	81	40	17	21	👥		Core	Schott (2005) ?Importers, Exporters, and Multinationals: A Portrait of Fir…
📝	5	81	40	15	22	👥	●	Base	Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. t…
📝	5	80	40	17	22	👥		Base	Wholesalers and Retailers in U.S. Trade (Long Version)
📝	5	80	40	15	23	👥	●	Core	Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. t…
📝	5	80	40	14	25	👥	●	Base	Firms in International Trade
📝	5	80	40	14	25	👥	●	Base	Firms in International Trade
📝	5	79	40	18	19	👥	●	Base	The Margins of U.S. Trade (Long Version)
📝	5	79	40	14	24	👥	●	Core	Firms in International Trade

Project ID 5: Impact of Foreign Trade on the U.S. Economy



Document Similarity Methods

- Keyword Search
  - Finding most prominent keywords in the abstracts of the research projects
  - Comparing to publication abstracts
- Natural Language Processing
  - Cosine Similarity
  - Understand the context of phrases and sentences in the research project
  - Scores the similarity of a given piece of text