Paper with Data

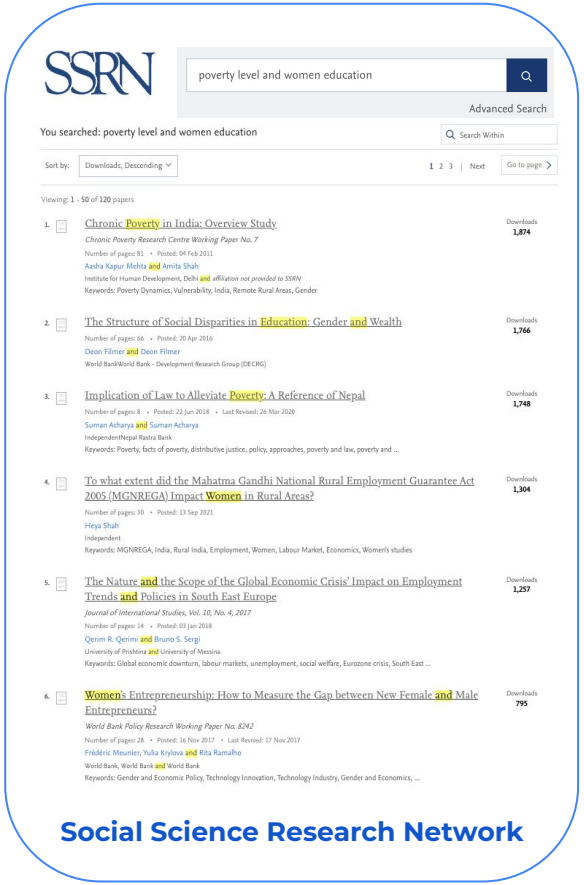
An Academic Search Engine through Fine-grained Dataset Information

AC215 Midterm Report

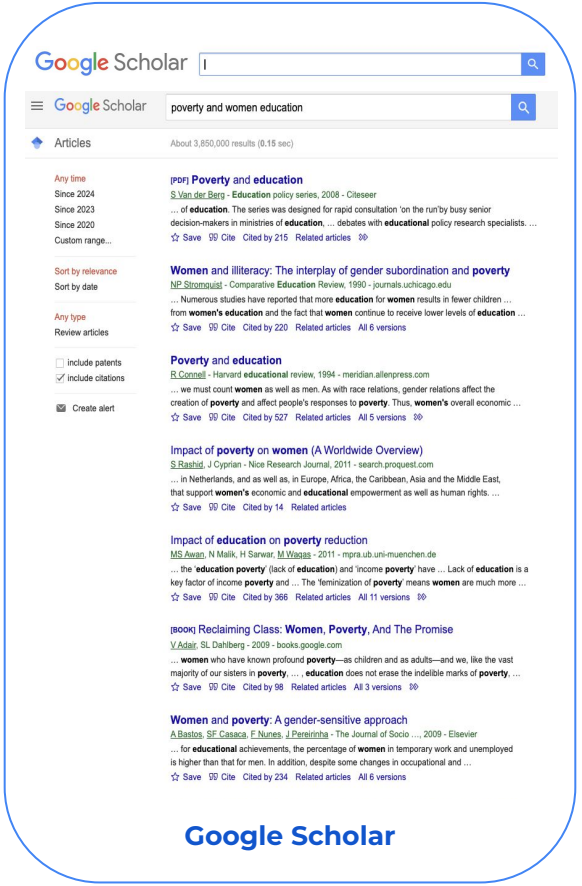
Team:
Zhiyu Li
Luozhong Zhou
Jinghan Huang
Haozhuo Yang

How long does it take to find the relevant research papers for your project?

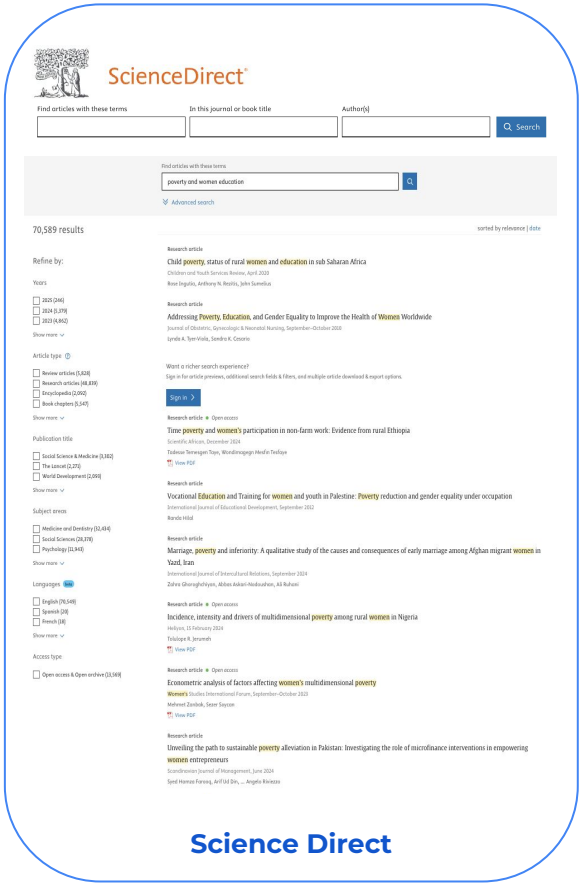
Existing Search Engine for Social Science Academia



Social Science Research Network



Google Scholar



Science Direct

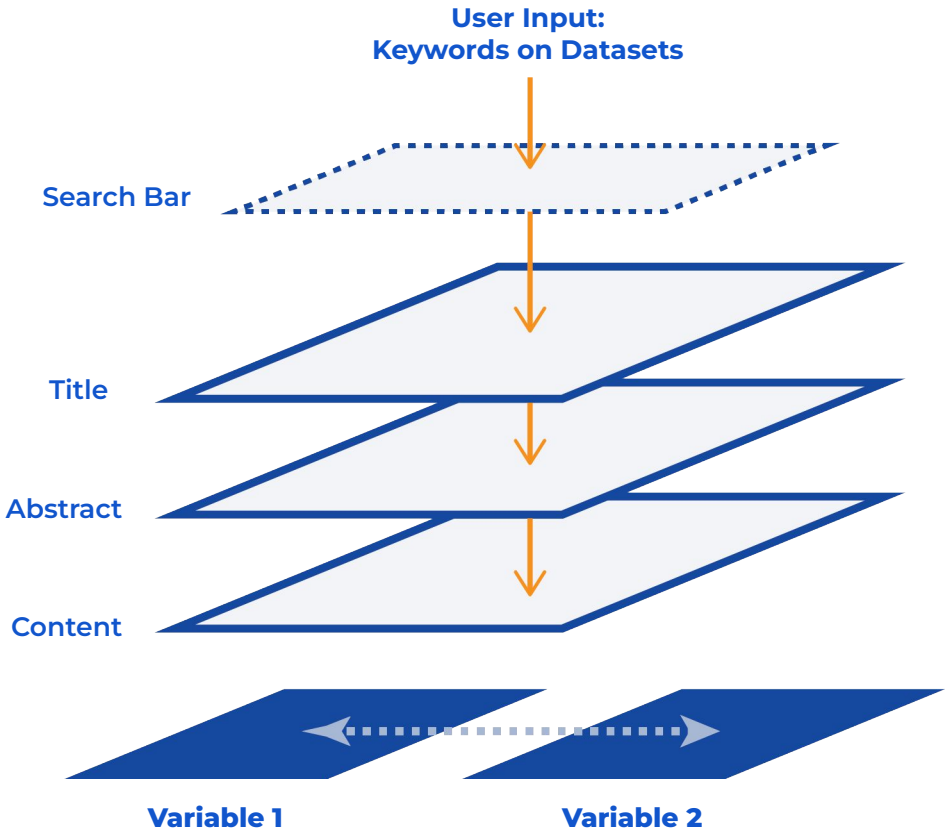
75 min

Average time for a scholar to find a relevant paper

<40%

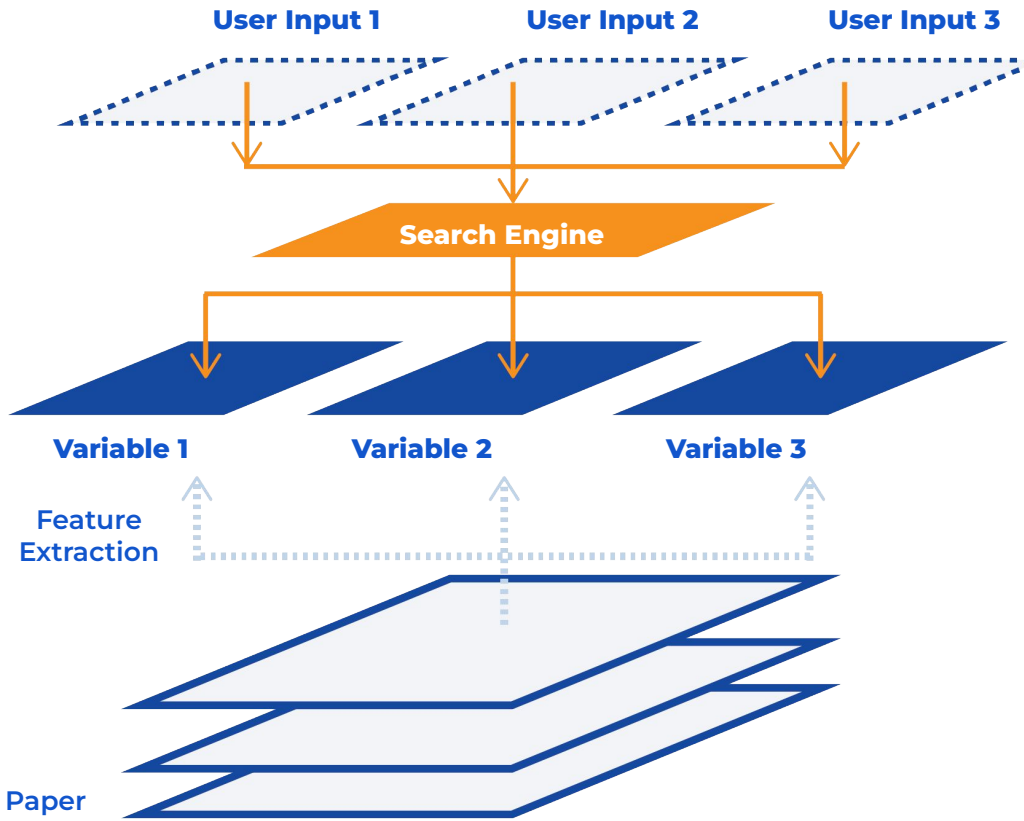
Of the search results are actually useful to the user

Problem Statement



Current social science academic search engines primarily rely on keyword matching within titles, abstracts, and content, often failing to address researchers' deeper needs for discovering relevant papers based on specific datasets and the relationships between variables.

A Fine-Grained Research Engine based on Datasets

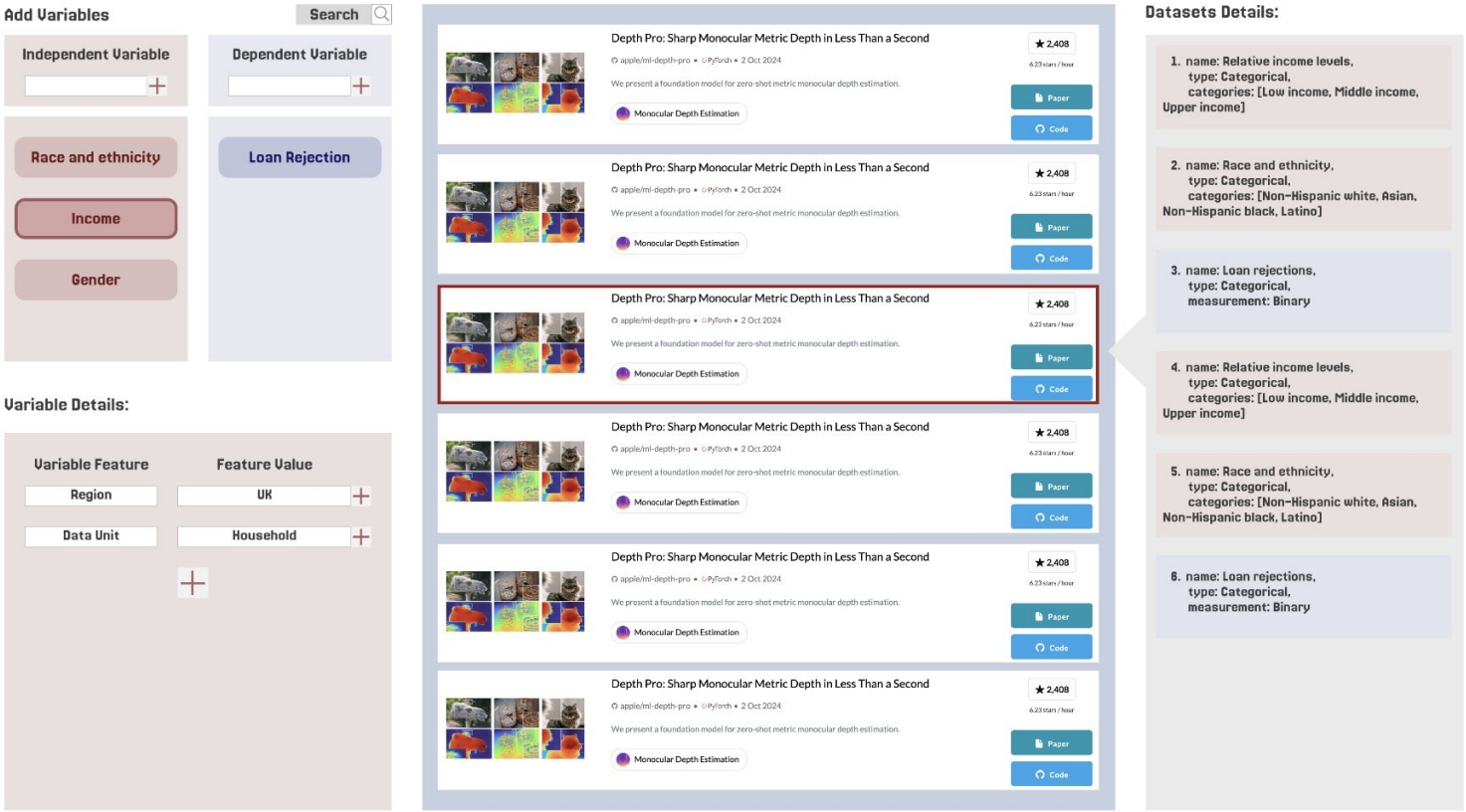


Our fine-grained engine searches papers by directly specifying data variables of interest.

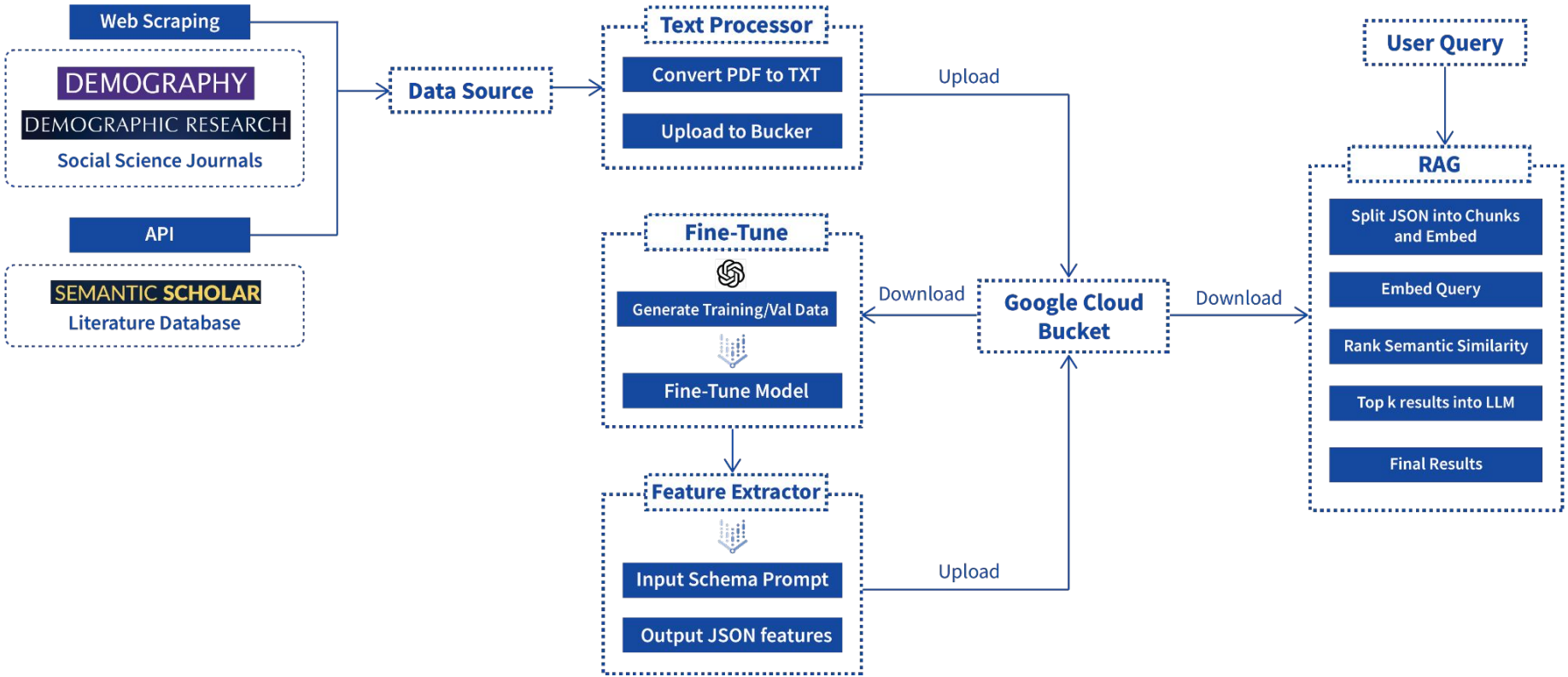
By extracting and indexing variables and their relationships from paper content, the system allows users to input multiple variable types as search criteria, delivering precise results that align with their specific data requirements. This approach streamlines the search process, helping researchers quickly identify studies that utilize the exact variables they need for their work.

User Interface

Paper with Data



Technical Architecture



Value Proposition

Enhanced Precision in Academia Searches

Unlike traditional keyword-based academic search engines, this platform offers fine-grained dataset categorization and relationship-based filtering. Researchers can specify dataset variables and relationships, quickly isolating relevant papers and reducing time spent filtering irrelevant results

Integration of Comprehensive Data Sources

Leveraging structured metadata from academic journals and popular datasets, this system provides an integrated search experience. With consistent data quality and full access to both metadata and article content, researchers benefit from a robust and reliable resource for literature discovery.

Future Development and Growth Potential

Diverse Input
Not just input variables, we want the user to directly input a prompt discussing their topics and what's most related to them. Then system returns the most relevant results based on the prompt.

Domain Expansion
Research domain doesn't need to be confined within social science. We will consider feature structures of different domains and support search for various scholars.

Collaboration and Sharing Features
Integrate social and collaborative features, allowing researchers to share curated lists, annotations, and search insights within research groups or communities.