

IssueLab Solves Data Puzzle to Empower Researchers

IssueLab and the Foundation Center provide free access to tens of thousands of case studies, evaluations, white papers, and issue briefs addressing some of the world's most pressing social problems. The amount of data made available by IssueLab grows every day, and they are tasked with making it easy for researchers to source relevant information needed to strengthen social sector initiatives. IssueLab tasked Datascope with enhancing the way information could be extracted and organized as searchable data, giving philanthropic innovators the ability to easily index and share the collective intelligence.

“Datascope moved quickly while still working collaboratively, which is a rare mix. Throughout design, iteration, and implementation, they ensured Sparck would integrate seamlessly into our system. The API gives us compelling new data-driven abilities that no off-the-shelf tools can provide.”

– Gabriela Fitz, Director of Knowledge Management Initiatives

Cataloging A Growing Library of Social Research

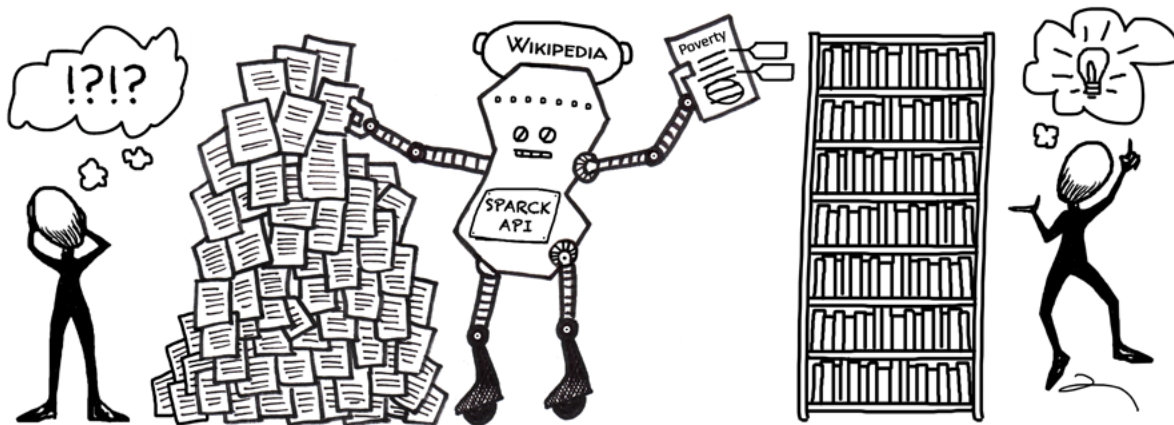
IssueLab, a service of the Foundation Center, works to collect research from the social sector in a centralized and open knowledge base. As their collection grows, IssueLab must introduce new ways to make relevant research easier to find and browse for their users.

IssueLab and Datascope tackled this problem by focusing on automatically extracting and creating metadata from the tens of thousands of social sector research documents in the IssueLab collection, with the knowledge that this metadata can enrich search, recommender systems, interactive data visualizations, and more.

The Sparck API: NLP Superpowers for IssueLab

Datascope began the project by experimenting with several techniques and worked in close collaboration with IssueLab to decide which were the most promising in terms of both initial results and value to the IssueLab platform.

After the prototype phase, Datascope designed and built the Sparck API (named for Karen Spärck Jones, a British pioneer in information retrieval), a robust RESTful API designed specifically to dovetail with IssueLab's



infrastructure. Among its features are advanced PDF text extraction, as well as other unique natural language processing tools designed to unearth relevant gems for researchers:

Key Findings: Many IssueLab documents have “key findings” sections with important takeaways for researchers. The Sparck API automatically extracts these when it can, and labels them for human review when it can't, saving IssueLab employees hours of time searching through documents for these nuggets of wisdom.

Wiktopic: The Sparck API includes a novel method of labeling documents with human-readable topics by identifying the Wikipedia articles and categories related to the language

in the document. These so-called “WikiTopics” make it easy for end users to quickly understand, at a glance, the relevant topics of 30-page documents, and find more context on the subject on Wikipedia.

Working Together to Enrich Collective Knowledge in the Social Sector

The collaboration between IssueLab and Datascope represents data science applied to some of the most crucial information for improving philanthropy and society as a whole. The modular and expandable Sparck API will be the engine behind myriad new enhancements and tools as IssueLab forges forward in their noble mission.

contact

information@datascopeanalytics.com if you have an inquiry.
www.datascopeanalytics.com

follow



meet

180 W. Adams, Suite 801
Chicago, IL 60603