

Problem

We've reached a tipping point where about half of the recent literature is open access, and this is growing quickly. At current rates of growth, all literature will be OA in the next few decades. And more realistically, OA mandates will lead to universal OA much earlier.

That's the good news. The bad news is that, although academic literature has never been more freely readable, it's far from being truly *accessible*. For regular people with an information need--whether it's a question about their diabetes medication, or about Japanese customs in the Edo period--the literature might as well be closed, because they don't always know where to find it. There's no easy way for laypeople to access the OA corpus.

Of course, they can use Wikipedia. But that's a secondary source, a summary. It's invaluable as a starting place, but it's not the knowledge itself. They can also use Google Scholar (GS). But there are a number of problems with GS:

1. Its interface is optimized for scholars, and intimidating to lay users.
2. It focuses on returning both OA and toll-access results. For lay users, it's vital to have the option to restrict search to *only* open access items.
3. It's a closed-data, closed-source side project run by a sprawling for-profit company. Google does not make any money from GS; they run it because they feel like it, for now. But Google have a long track record of killing much-loved products they've lost interest in, including Reader, Buzz, Answers, Health, Wave, Code, and dozens more. Many of these were more lucrative and popular than GS. From this perspective, it seems only a matter of time before Google kills Scholar as well. And because GS data and code are both closed, there's no chance for the community to pick up support.

Solution

We propose a search engine that helps laypeople discover and read open access papers and books. Here are some of the features of this search engine:

- It has a simple, clear interface built to be easy for laypeople to use, including K12 students.
- It allows users to search *only* open access resources by default.
- It uses machine learning to cluster search results in an intelligent way. For example, a user searching for Edo period Japan would see a suggestion to include "Tokugawa period" as well, since this is another name for Edo period. This and similar clusters can be discovered in the background by automatic mining of the open access fulltext in the system. It's important because, unlike specialists, lay users typically begin searches with little knowledge of field-specific jargon.

- It's optimized for use on mobile phones, helping to support use in the developing world where most internet use is on mobile devices. It's also available in languages other than English.

Why us?

Although we're a small organization, Impactstory has a five-year history as an energetic, sustainable, open nonprofit. We think we're a great team to make this project a reality.

- We've already got the data. Our oaDOI system has 18M fulltext open access scholarly resources (papers, book, proceedings, and more) with DOIs, right now. This is the largest collection of its kind in the world.
- We know how to build at scale. oaDOI is used by thousands of academic libraries worldwide, via integration with Web of Science and the SFX link resolver. We handle more than one million API calls per day.
- We know how to build simple, powerful UIs for a lay audience and then reach that audience online. Our Unpaywall browser extension is used by over 100,000 users, and has been featured in outlets like TechCrunch, Gizmodo, and Mental Floss.
- We're good value: we've done this work (as well as projects like Impactstory Profiles, Depsy, CiteAs, and more) using a "fast and light" strategy relying on a small, agile, passionate team. This allows us to create great tools quickly and very inexpensively, reducing funder risk.
- We're deeply engaged in the OA and Open Science communities.

Here's the point

The Library of Alexandria, while it stood, stored vast quantities of humankind's accumulated knowledge. Its collection was built at tremendous cost and effort (including, reportedly, the summary seizure of any books found on ships entering Alexandria's harbour). The result was a monument to academic learning--but a hollow one, since few were permitted to actually *access* the library's content.

Today, humankind has assembled a new treasure of knowledge, an increasingly open corpus of research we often describe as "the scholarly record." But like the library at Alexandria, it has remained closed to most of humanity--first because of subscription tolls, and now because of the dearth of OA search tools. Obscurity is the new gatekeeper.

We can smash the gate. By creating a simple, open search system, we create a new, open, virtual Library of Alexandria: OpenAlex. From the outset it will move millions of scholarly records into public accessibility. And as OA becomes increasingly universal, OpenAlex will become an indispensable treasury of human knowledge--a monument to the idea that our species can learn, together. In today's world, this has never been more important.

