Richtlinie zum „Software-Sprint“

InvestigatingScience

Werkzeuge zur Untersuchung von Forschungsintegrität

Schlussbericht

Zuwendungsempfänger:

Hristio Boytchev

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS24S23 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Aufgabenstellung und Motivation

Probleme mit Forschungsintegrität sind verbreitet und schaden dem Wissenschaftssystem. Kürzlich sind etwa hochrangige und für das jeweilige Feld wegweisende Veröffentlichungen über Alzheimer unter Fälschungsverdacht geraten.

Gleichzeitig sind die Kontroll- und Korrekturmechanismen des Wissenschaftssystems teils ungenügend. So werden selbst nachgewiesen mangelhafte Studien nicht zuverlässig zurückgezogen und wenn dies geschieht, werden sie teilweise trotzdem weiter zitiert, bleiben also Teil des Wissenskorpus. Überführte Forschende erfahren höchst selten Konsequenzen.

Wir wollen mit unserem Projekt „Investigating Science" kritischen Wissenschaftsjournalismus durch die Bereitstellung von datenbasierten Recherchemethoden stärken. Wir nutzen dafür Bibliometrie, die Auswertung von wissenschaftlichen Publikationsdatenbanken, um Hinweise auf problematische Praktiken zu identifizieren. Wir haben uns auf problematische Forschungsförderung, nicht begutachtete Pre-Prints und Predatory Journals (also wissenschaftliche Journale ohne Qualitätskontrolle) konzentriert.

Dafür haben wir Beispielanalysen, Rezepte und Code zum eigenen Durchführen dieser Analysen entwickelt.

Folgende Meilensteine waren benannt:

1) Definition der bibliometrischen Analysemethoden

- Vorrecherche zu bereits veröffentlichten bibliometrischen Methoden
- Vergleich und Festlegung auf mindestens drei Fragestellungen (etwa Retractions, Predatory Journals, Zitationskartelle)
- Auswahl der Datenquellen

- Dokumentation und Aufbereitung der Skripte

2) Skalierung durch den Aufbau einer Rechercheplattform

- Prüfung von Ansätzen, um über eine Online-Plattform rechtlich und technisch machbar direkte Abfragen (etwa zu Instituten, Wissenschaftler:innen, Journalen sowie deren Verknüpfungen) an Endnutzer:innen bereitzustellen. Als mögliche Datenquellen dienen bibliometrische Datenbanken wie openalex.org, doaj.org, semanticscholar und openaire, zudem die Follow the Grant Rechercheplattfor.

3) Entwicklung des Frontends der Plattform

- Veröffentlichung der Rechercherezepte, inklusive Code und einer Anleitung zur eigenen Durchführung
- Optionaler Zugang samt Suchmaske (nach Wissenschaftler:innen, Institutionen, etc.)zu direkten Datenquellen
- Visualisierung der Analyseergebnisse

4) Kommunikation nach außen

- Publikation einer Beispielrecherche mithilfe der Methoden und Daten
- Vortrag auf einer Wissenschafts- oder Journalismuskonferenz

Beitrag des Projektes zu den Zielen der Förderinitiative „Software-Sprint“

Die Arbeit trägt zu „Data Literacy“ der journalistischen Zielgruppe und dann über deren Arbeit auch zum Teil der Öffentlichkeit bei. Als Projekt, das kritischen Journalismus mit Software- und Datenmethoden stärkt, kann es zudem als „Civic Tech“ gewertet werden.

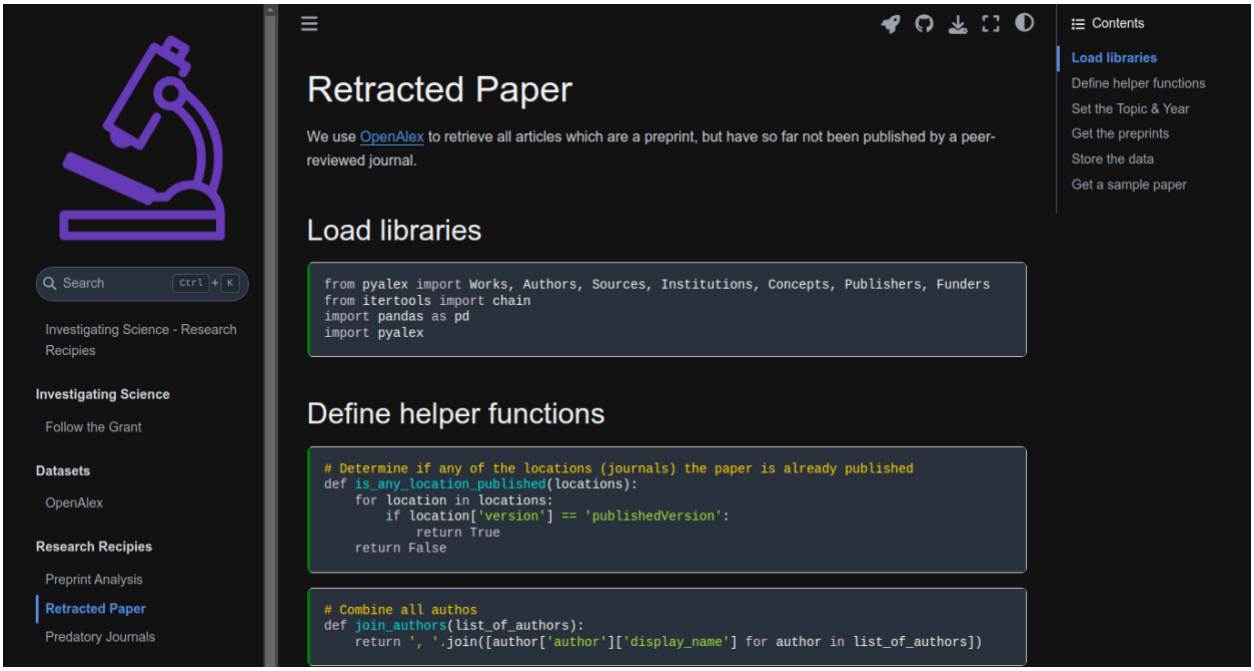
Ausführliche Darstellung der Ergebnisse

Wir haben innerhalb der Förderzeit erfolgreich unsere Ziele und Meilensteine erreicht.

Durch eine Vorrecherche und Machbarkeitsanalyse haben wir zunächst die Ansätze Forschungsförderung, nicht begutachtete Pre-Prints und Predatory Journals für die Ausarbeitung ausgewählt.

Nach einer Recherche unterschiedlicher Datenquellen haben wir OpenAlex als den am meisten versprechenden Ansatz gewählt, da es eine offene Datenbank ist, die gegenwärtig auch rasant Verbesserungen erfährt.

Anschließend haben wir die Recherchemethoden ausgearbeitet, dokumentiert und samt einfachen Visualisierungen veröffentlicht.



Beispielskript

Auf die als optional benannte direkte Bereitstellung von Daten wurde vor allem aufgrund rechtlicher Überlegungen verzichtet.

Das Projekt wurde anschließend auf der Datenjournalismus-Konferenz Scicar vorgestellt. Das Feedback war durchweg positiv und es kam zu Anfragen über Zusammenarbeit bzw. Nutzung der Methoden.

Eine erste journalistische Recherche ist vorangeschritten und wird voraussichtlich bald erscheinen . Hier ist es zu einer Verzögerung gegenüber dem Antrag gekommen.

Zusätzlich haben wir einige Verbesserungsmöglichkeiten in der Datenbank OpenAlex herausgearbeitet, die wir der Organisation zurückgespiegelt haben.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Wir verfolgen als Projekt das Ziel, Journalist:innen zu mehr kritischer Recherche über Wissenschaft zu ermächtigen. Innerhalb dieser Förderung haben wir dazu einen wichtigen Schritt gemacht. Weitere Früchte aus dieser grundlegenden Arbeit erhoffen wir uns auch für die Zukunft. Wir sind dabei, das Projekt mit dem Vorgänger “Follow the Grant” zusammenzuführen und so eine Nachhaltigkeit sicherzustellen.

Persönlich hat mir die Arbeit mir wichtige Erfahrung mit datenbasierten Recherchen gebracht. Ich hatte den Raum, unterschiedliche Ansätze auszuprobieren und die vielversprechenden weiterzuverfolgen. Es hat sich aber gezeigt, dass Ergebnisse, die einer journalistischen Veröffentlichung dienen, eine Präzision erfordern, die von Datenquellen und auch automatisierten Methoden nicht sichergestellt werden kann. Im Anschluss bedeutet das, dass eine manuelle Überprüfung notwendig ist, die den Arbeitsaufwand erheblich und teils restriktiv erhöhen kann.

Darüber hinaus hat mir das Projekt auch durch die Vernetzung mit der Kohorte und das angebotene Coaching wichtige Erkenntnisse über Softwareentwicklung und Projektmanagement gebracht.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Wir haben uns aufgrund rechtlicher Bedenken gegen eine direkte Veröffentlichung von Daten (etwa Namen von Instituten, Wissenschaftler:innen und Journalen) entschieden. Stattdessen ist es möglich, durch unsere Skripte und offene Datenquellen zu solchen Daten zu gelangen.

Der Ansatz, quantifizierbare Ergebnisse direkt aus OpenAlex zu erhalten, war mit sehr vielen Herausforderungen verknüpft. Oftmals waren die Metadaten zu den wissenschaftlichen Artikeln unvollständig oder inkorrekt. Dies führte zu falschen Verknüpfungen zwischen Instituten, Firmen und Geldgebern.

Die Gründe dafür sind vielfältig: Einerseits ist die automatische Textextraktion noch nicht ausgereift. Andererseits sind die Formulierungen zur finanziellen Unterstützung von wissenschaftlichen Arbeiten oft sehr vage formuliert.

Wir glauben trotzdem, dass OpenAlex ein guter Ansatz für investigativen Datenjournalismus ist, räumen allerdings der manuellen Verifikation der Daten mehr Zeit ein und geben dieses Wissen auch an andere Journalist:innen weiter.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Hier gibt es Informationen zum Projekt:

- Workbook: https://investigatingscience.github.io/research_recipies_workbook/basics/intro.html#
- Webseite mit Beispielrecherchen: https://followthegrant.org/
- GitHub: https://github.com/InvestigatingScience/research_recipies_workbook

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Die Kostenplanung wurde eingehalten.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Für die Veröffentlichungen der Recherchen sind wir auf journalistische Medien und deren Zeitplan angewiesen. Dadurch ist es zu einer Verzögerung einer ersten Veröffentlichung über den Förderzeitraum hinaus gekommen.