

KI- und zitationsbasierte Tools für die Literatursuche

AI- and citation-based tools for literature search

Abstract

Various newer tools use AI and/or citation-based algorithms to display a small number of “good” hits. You can either start with a specific paper (Connected Papers, ResearchRabbit) or you can enter a question (SciSpace, Elicit). With the latter tools, papers are summarised and data can be extracted, even from uploaded PDFs. The tools are briefly introduced and compared with each other.

Keywords: literature search, AI tools, citation tracking, data extraction

Zusammenfassung

Verschiedene neuere Tools nutzen KI und/oder zitationsbasierte Algorithmen, um wenige „gute“ Treffer anzuzeigen. Man beginnt entweder mit einem bestimmten Paper (Connected Papers, ResearchRabbit) oder kann eine Frage eingeben (SciSpace, Elicit). Bei den letzteren Tools werden Paper zusammengefasst und können Daten extrahiert werden, auch von hochgeladenen PDFs. Die Tools werden kurz vorgestellt und miteinander verglichen.

Schlüsselwörter: Literatursuche, KI-Tools, Citation Tracking, Datenextraktion

Einleitung

Mit Aufkommen von Large Language Models (LLMs) wie vor allem GPT beschäftigen sich Bibliotheken auf Tagungen und Veranstaltungen oder über Zeitschriften, Blogs und Social Media mit neueren Tools zur Literatursuche jenseits von etablierten Artikeldatenbanken wie PubMed, Embase und Web of Science.

Die Beurteilung dieser Tools stellt Bibliothekar*innen vor einige Herausforderungen. Natürlich wollen wir die Hauptfunktionen verstehen, die im Folgenden für einige ausgewählte Tools beschrieben werden, aber wir sollten auch die Frage beantworten können, was die Anwendungszwecke sind, also konkret: für welche Situationen würden wir Studierenden/Ärzt*innen/Forschenden die Benutzung empfehlen?

Der Markt ist sehr schnell gewachsen, inzwischen gibt es einige Online-Tools mit teils ähnlichen Funktionen (siehe z.B. <https://confluence.frankfurt-university.de/pages/viewpage.action?pageId=225216646>). Exemplarisch werden hier die beiden Tools für Literatursuche und Datenextraktion SciSpace und Elicit vorgestellt, die explizit mit Artificial Intelligence (AI) werben. Andere neuere Tools wie Connected Papers und ResearchRabbit nutzen vorrangig Zitationsnetzwerke mit Visualisierung, ohne KI/AI oder LLMs zu erwähnen, arbeiten also mit bestimmten Algorithmen, und verlangen am Anfang die Auswahl eines bestimmten Papers.

Volker Braun¹

¹ Universitätsmedizin
Mannheim, Medizinische
Fakultät Mannheim der
Universität Heidelberg,
Mannheim, Deutschland

Citation Tracking

Citation Tracking

Für das Verständnis von zitationsbasierten Tools ist ein Überblick über verschiedene Citation-Tracking-Methoden hilfreich [1]. Man kann zwischen direct (forward, backward) und indirect citation tracking (co-cited bzw. co-citation, co-citing) unterscheiden. Eine gute Erläuterung liefern Klerings et al. im Appendix C von [2]: „Citation-based search methods include: backward citation searches (i.e., reference list checking) and forward citation searches (i.e. „cited by” searching); co-cited references (i.e., articles that are cited together with the seed article) and co-citing references (i.e., articles that share similar reference lists with the seed article).“ Co-citing wird auch bibliographic coupling genannt [3]. Das Konzept ist schon sehr lange in Web of Science als Funktion „Related records“ verfügbar, sortiert nach der Anzahl der shared references. Das Konzept co-cited/co-citation, also andere references werden *mit* dem Ursprungspaper („seed article/paper“) häufiger zitiert, wird aktuell m.W. nur von Connected Papers genutzt. Citationchaser ist eine kostenfreie Alternative zu z.B. Web of Science für die Anzeige von references (backward) und citations (forward).

Connected Papers und ResearchRabbit

Connected Papers

Connected Papers' „similarity metric is based on the concepts of *Co-citation* and *Bibliographic Coupling*“ [4]. Laut Aussage eines Co-Founders per E-Mail ist der Algorithmus „mostly based on graph theory“. Wie von anderen Tools auch wird als Datengrundlage Semantic Scholar genutzt. 5 „graphs“ pro Monat sind nach Registrierung frei, kostenpflichtige Academic- und Business-Optionen sind ab rund 4,50 Euro lizenzierbar.

Nach Auswahl eines Papers („Search by keywords, paper title, DOI or another identifier“) werden 40 ähnliche Paper als „graph“ (Netz mit Punkten) oder Liste angezeigt, „Prior works“ („These are papers that were most commonly cited by the papers in the graph.“) und „Derivative works“ („These are papers that cited many of the papers in the graph.“) gibt es als zusätzliche Optionen. Für Literaturverwaltungsprogramme können .bib-Dateien (BibTeX) heruntergeladen werden.

ResearchRabbit

ResearchRabbit (<https://researchrabbitapp.com/>) ist komplett kostenfrei, man muss aber bei der Registrierung mehr Daten angeben. Die Artikeldaten kommen laut FAQ von Semantic Scholar, OpenAlex und PubMed. Man startet mit einer selbst angelegten Collection und dann der Auswahl eines Papers, auch via PMID. Dann werden Paper als „Similar Work“ angezeigt (anfangs 50, „Load More“ möglich), als Liste und als „graph“ (Netz mit Punkten), plus „References“ und „Citations“ des Ursprungspapers, nach Hinzufügen von Papern zur Collection wird das für mehrere ausgewählte Paper zu „Earlier Work“ und „Later Work“ umbenannt. Der Support schreibt auf Anfrage, dass „co-cited references and co-citing references as part of the recommendation algorithm“ genutzt würden. Für Literaturverwaltungsprogramme können .bib- (BibTeX) oder RIS-Dateien heruntergeladen werden.

Vergleich

Bei einem exemplarischen Vergleich von 41 Treffern von Connected Papers (CP, mit seed paper) bzw. 50 von ResearchRabbit (RR) mit den „cited“/„citations of“/„related records“ (top 50) von der Web of Science Core Collection (WoS CC, insgesamt 216) bezogen auf ein bestimmtes Paper war das Ergebnis: CP: 27 von 41 sind Dubletten, RR: 22 von 50 sind Dubletten – es war zu erwarten, dass es einen relevanten Anteil an Überschneidung gibt. Allerdings gab es kaum Überschneidungen zwischen CP und RR, nur 3 Dubletten, hier scheint also die Funktionsweise ziemlich unterschiedlich zu sein.

Ein Vergleich mit nur den „related records“ von WoS CC wäre noch denkbar, ebenso der zusätzliche Vergleich mit den top 50 „Similar articles“ aus PubMed (siehe <https://pubmed.ncbi.nlm.nih.gov/help/-#similar-articles> für Details).

SciSpace und Elicit

SciSpace

SciSpace (via <https://typeset.io>) arbeitet in der Funktion „Literature Review“ mit freisprachlicher Suche, Formulierungsvorschlägen, einer Zusammenfassung von „top 5 papers“ (10 kostenpflichtig, ebenso „High quality“ als besseres Sprachmodell) und darunter 10 Treffern in einer Tabelle („Load more“ möglich) mit Zusammenfassungen (und ggf. Erläuterungen bei geringer Literaturauswahl wie z.B. „The paper does not specifically address ... However ...“) und der Möglichkeit, in der freien Version fünf zusätzliche inhaltliche Aspekte als vordefinierte oder selbst formulierte Spalten zu ergänzen, z.B. „Conclusions“ (50 Spalten mit dem kostenpflichtigen Premium-Account), siehe Abbildung 1. Nach Anhaken hat man über „Select an action“ mehrere Optionen, z.B. „Show more like selected“, siehe Abbildung 2. Als Datenquellen wurden vom Support im November 2024 „OpenAlex and MAG. Sometimes from Semantic Scholar“ genannt (MAG bedeutet Microsoft Academic Graph; Support: „Microsoft retired MAG in 2021, but it served as a valuable tool for understanding research trends, relationships, and academic impact.“). Ein Datenexport ist nur ab Premium möglich. Man kann via „Chat with PDF/Paper“ eigene oder vorausgewählte Fragen an ein hochgeladenes oder als Volltext verfügbares Paper stellen und sich markierte Textstellen erklären lassen. Absätze werden bei Mouseover automatisch markiert, allerdings ist das Markieren mehrerer Absätze schwierig.

Weitere Funktionen sind u.a. „Extract Data: Get summary, conclusions and findings from multiple PDFs in a table.“, „Find Topics“, „AI Writer“, „Citation Generator“ und „Paraphraser“, das Tool ist also sehr umfangreich. Zudem gibt es Lizenzen für „Labs & Universities“ mit Kosten pro User (je mehr, desto günstiger). Bei Support-Bedarf kann man mit Chatbots chatten oder eine E-Mail schreiben.

Elicit

Elicit (<https://elicit.com/>) funktioniert ähnlich wie SciSpace, ist jedoch etwas eingeschränkter. Via „Find papers“ ist freisprachliche Suche (natural language) möglich, daraufhin wird automatisch ein Titel als „Notebook“ angelegt und als Ergebnis wird eine „Summary of top 4 papers“ (mit dem kostenpflichtigen Plus-Account 8) angezeigt und darunter 8 Treffer in einer Tabelle („Load more“ möglich) mit einsätzigen Zusammenfassungen und der Möglichkeit, in der freien Version zwei zusätzliche Aspekte als Spalten zu ergänzen, z.B. „Methodology“ und

Abbildung 1: SciSpace – Beispiel Literatursuche

Abbildung 2: SciSpace – „Select an action“

„Main findings“ (5 zusätzliche Spalten mit Plus). Zur Kontrolle zeigt ein Klick in eine Zelle relevante (hervorgehobene) Textstellen aus Abstract oder Volltext an. Was davon zur Analyse verfügbar ist, wird in der Tabelle nicht angezeigt. Plus bietet nur „1 high-accuracy column per table“, bei dem ein besseres Sprachmodell eingesetzt wird, im Pro-Abo unlimitiert. Zudem können 10 PDFs pro Monat zur Analyse hochgeladen werden („Extract data from PDFs“), kostenpflichtig 25 oder mehr. PDFs können selbst hochgeladen werden und sind dann in der „Library“ oder im Reiter „Extract data from PDFs“ verfügbar. Ein Datenexport ist nur ab Plus möglich.

„Show more like these“ nach Auswahl bedeutet, dass „references“ and „cited by“ papers angezeigt werden. Ebenfalls nach Auswahl gibt es via „Add a new step“ die Beta-Funktion „Chat with papers“, allerdings bei „Use full text“ mit dem Hinweis „Upgrade to Plus to chat with full papers“, das aber ohne Tabellen („excluding tables“). Ein

Chat mit Abstracts in der freien Version erscheint mir wenig sinnvoll.

Zudem gibt es noch die Funktion „List of concepts“. Nach z.B. der Eingabe von „treatments for obesity“ wird automatisch das Notebook „Innovative Approaches to Obesity Treatment“ angelegt und eine Tabelle mit den Spalten Concept und Source und den darüber stehenden Hinweisen als Liste „Found 72 papers, Found 383 concepts in 72 papers, Found 115 unique concepts, Final answer“ (Stand 4.11.2024).

Datenextraktion

Eine Arbeitsgruppe der Fakultät hat im April 2024 die Datenextraktion anhand eines Papers mit SciSpace Basic und Premium auf meine Empfehlung hin getestet, ich habe das Gleiche mit Elicit Basic durchgeführt. Es gab bei Elicit Probleme mit Informationen aus Tabellen/Gra-

fiken (für Tabellenanalyse ist Plus nötig). Eine Auswahlmöglichkeit bei Antworten ist bei beiden Tools nicht möglich, d.h. jede mögliche Frage muss einzeln formuliert werden. Das Hauptproblem war aber: Antworten waren sowohl in Elicit Basic auch mit SciSpace Basic und Premium öfter falsch. Elicit Plus müsste allerdings noch getestet werden. Vermutlich sollte man aufgrund des Entwicklungsfortschritts der Tools solche Tests regelmäßig durchführen.

Fazit

Generell untersuchungswürdig bei SciSpace und Elicit ist die Verwendung eines besseren Sprachmodells nur gegen Bezahlung. Da stellt sich die Frage, ob die erzeugten Daten beim Standardmodell falsch sein könnten.

Bezogen auf „citation searching“ empfiehlt TARCIS „For systematic search topics that are difficult to search for, backward and forward citation searching should be seriously considered as supplementary search techniques.“ und „... should be based on all included records of the primary search“ [5]), siehe auch <https://tarcis.unibas.ch/>). Das ist eine eng gefasste Möglichkeit der Nutzung, dafür würde auch Web of Science Core Collection, Scopus oder Citationchaser oder eine ähnliche Datenbank ausreichen.

Denkbar ist die Nutzung von Connected Papers, ResearchRabbit, Elicit und SciSpace meines Erachtens für die Gewinnung von mehreren Papern, die man für die Generierung von Suchbegriffen (Analyse von Titel, Abstract, ggf. Schlagwörtern) und die Verifizierung von Suchstrategien verwendet (diese Artikel müssen gefunden werden). Oder es werden für kleinere Projekte von Studierenden, z.B. zur Postererstellung, nur wenige passende Nachweise benötigt. Das konkurriert dann ggf. mit schon lange verfügbaren Funktionen wie z.B. „Similar articles“ in PubMed. Die Einarbeitung in neue Themengebiete mit wenigen guten Publikationen ist auch möglich, allerdings gibt es auch dafür alternative Datenbanken wie DynaMed, Trip oder ACCESSSS.

Eine Verwendung für breitere Suchen erscheint mir aufgrund der unterschiedlichen Funktionsweisen im Vergleich zur herkömmlichen (Booleschen) Datenbanksuche in z.B. PubMed nicht angebracht. Allerdings sollte das mit geeigneten Methoden wissenschaftlich untersucht werden.

Für Datenextraktion konnten die beiden exemplarisch getesteten Tools zum Zeitpunkt des Tests nicht überzeugen, das müsste aber breiter und neu geprüft werden.

Anmerkung

Interessenkonflikte

Der Autor erklärt, dass er keine Interessenkonflikte in Zusammenhang mit diesem Artikel hat.

Literatur

1. Hirt J, Nordhausen T, Appenzeller-Herzog C, Ewald H. Using citation tracking for systematic literature searching – study protocol for a scoping review of methodological studies and a Delphi study. *F1000Res*. 2020 Dec 01;9:1386. DOI: 10.12688/f1000research.27337.3
2. Klerings I, Robalino S, Booth A, Escobar-Liquitay CM, Sommer I, Gartlehner G, Devane D, Waffenschmidt S; Cochrane Rapid Reviews Methods Group. Rapid reviews methods series: Guidance on literature search. *BMJ Evid Based Med*. 2023 Nov 22;28(6):412-7. DOI: 10.1136/bmjebm-2022-112079
3. Kessler MM. Bibliographic Coupling between Scientific Papers. *American Documentation*. 1963 Jan;14(1):10-25.
4. Connected Papers. About. [letzter Zugriff am 26.11.2024]. Verfügbar unter: <https://www.connectedpapers.com/about>
5. Hirt J, Nordhausen T, Fuerst T, Ewald H, Appenzeller-Herzog C; TARCIS study group. Guidance on terminology, application, and reporting of citation searching: the TARCIS statement. *BMJ*. 2024 May;9(385):e078384. DOI: 10.1136/bmj-2023-078384

Korrespondenzadresse:

Volker Braun

Bibliothek, Medizinische Fakultät Mannheim der Universität Heidelberg, Universitätsklinikum, Haus 42, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Deutschland
volker.braun@medma.uni-heidelberg.de

Bitte zitieren als

Braun V. KI- und zitationsbasierte Tools für die Literatursuche. *GMS Med Bibl Inf*. 2024;24(2):Doc28. DOI: 10.3205/mbi000611, URN: urn:nbn:de:O183-mbi0006117

Artikel online frei zugänglich unter

<https://doi.org/10.3205/mbi000611>

Veröffentlicht: 18.12.2024

Copyright

©2024 Braun. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.