

Einige K.I.-unterstützte Anwendungen für die Wissenschaft und Forschung

Grundsätzliches: Die Auswahl an K.I.-unterstützten Anwendungen im Internet wird zunehmend unüberschaubar und immer mehr Start-Ups versuchen ihr Glück mit ihrem eigenen Angebot. Allen Alternativen gemein ist dabei die Kombination aus einer mit Metadaten versehenen Literaturdatenbank mit enthaltenen Volltexten und einem Large Language Model (LLM) – für gewöhnlich ist das GPT-3, Elicit nutzt jedoch eine BERT-Abwandlung. Die LLMs erhalten ein Finetuning auf Basis der wissenschaftlichen Texte und erfüllen in der Regel eine der drei Funktionen:

- a) Sie synthetisieren die für eine Suchanfrage relevantesten Artikel, welche in der Datenbank vorhanden sind, und generieren einen entsprechenden Text (quasi ein Mini Literature Review)
- b) Sie extrahieren relevante Textpassagen aus den gefundenen Artikeln und geben diese aus, statt selbst Text zu generieren
- c) Sie ermöglichen eine «semantische» Suche, d.h. sie berücksichtigen Synonyme bei der Suchanfrage. (Bsp: eine Suche nach «anxiety» liefert auch Artikel über «panic attacks»)

Im Folgenden erhalten Sie einen Überblick über drei ausgewählte K.I.-unterstützte Anwendungen für die Wissenschaft und Forschung. Andere Angebote werden im Grossen und Ganzen ähnliche Funktionen haben.

Scite.ai

- Enthält 25 Millionen wissenschaftliche Publikationen aus den folgenden Datenbanken: <https://help.scite.ai/en-us/article/where-do-you-get-your-articles-from-1vglydm/>
 - Die Scite.ai-Datenbank enthält zudem 1 Milliarde citation statements, die aus den Publikationen extrahiert wurden
 - Def. "Citation statement": "sentences surrounding the citation within a document" (Nakov/Schwartz/Hearst 2004)
- Fachlicher Fokus liegt auf den Lebens- und Naturwissenschaften
 - Schreibstil in diesen Disziplinen ermöglicht insbesondere die automatisierte Identifikation und Klassifikation von citation statements»
- Arbeitet mit dem «smart citation index»:
«Smart Citations reveal how a scientific paper has been cited by providing the context of the citation and a classification system describing whether it provides supporting or contrasting evidence for the cited claim, or if it just mentions it" (Nicholson et al. 2021, 883)
- Smart citations enthalten folgende Informationen
 - a. Citation statement/citance
 - b. Citation context
 - c. Location of citation in the chapters
 - d. Intent (support, contrast, mention)
 - e. Editorial information (Nicholson et al. 2021, 883)
- Scite-ai setzt auf das LLM SciBert (fine-tuned) (Nicholson et al. 2021, 890)
- Funktion des LLM:
 - d) Das LLM versucht, ein gegebenes citation statement zu verstehen und in der Datenbank andere statements zu finden und auszugeben, welche die Aussage
 - a) unterstützen
 - b) widersprechen
 - c) erwähnen (<https://help.scite.ai/en-us/article/how-are-citations-classified-1a9j78t/>)

Elicit.com

- Greift auf die Literaturdatenbank von Semantic Scholar zurück
- e) Semantic Scholar enthält öffentlich zugängliche Aufsätze, vorrangig aus der Biomedizin (AI2 2018, 145f.)
- f) Semantic Scholar ist selbst K.I.-unterstützt und filtert Suchergebnisse auf ihre unmittelbare inhaltliche Relevanz hin → Semantic Scholar gibt in der Regel nur wenige Suchergebnisse aus (↔ Google Scholar; cf. AI2 2018, 145)
- Ermöglicht eine promptbasierte Suche nach Literatur
- g) Wie Scite.ai ist Elicit vor allem auf Lebens- und Naturwissenschaften spezialisiert
- Setzt auf GPT-3 (<https://musingsaboutlibrarianship.blogspot.com/2022/11/g-academic-systems-elicitorg-scispace.html>)
- Funktion des LLM:
 - h) Ermöglicht eine semantische Suche
 - i) Gefundene Artikel werden durch einen generierten Text zusammengefasst
 - j) Die Ergebnisse der relevantesten Ergebnisse werden synthetisiert und in Form eines «Mini-Literature-Reviews» ausgegeben (K.I.-generierter Text!) (<https://musingsaboutlibrarianship.blogspot.com/2022/11/g-academic-systems-elicitorg-scispace.htm>)
- Auf der Seite angegeben: accuracy of 90% of the answers given (<https://elicit.com/>)

Consensus.app

- Nutzt massgeblich die Literatur-Datenbank von Semantic Scholar (siehe Elicit.com)
- Funktion des LLM:
 - Relevante Textausschnitte werden extrahiert und angezeigt. Im Unterschied zu Elicit werden also keine Texte von der K.I. generiert!
 - ⇒ Consensus gibt zitierfähiges Material aus

Quellen

Allen Institute for Artificial Intelligence (AI2). (2018). Semantic Scholar. *Journal of the Medical Library Association*, 106 (1), 145-147.

Nakov, P., Schwartz, A., & Hearst, M. (2004). Citances. Citation Sentences for Semantic Analysis of Bioscience Text. In *Proceedings of the SIGIR'04 workshop on Search and Discovery in Bioinformatics*.

Nicholson, J., et al. (2021). Scite. A Smart Citation Index that Displays the Context of Citations and Classifies Their Intent Using Deep Learning. *Quantitative Science Studies*, 2 (3), 882-898.