Simplify your Search Engine Research

Wie das Result Assessment Tool (RAT) Studien auf der Basis von Suchergebnissen unterstützt

Sebastian Schultheiß[1], Sebastian Sünkler[1], Nurce Yagci[1], Daniela Sygulla[1], Sonja von Mach[1], Dirk Lewandowski[1,2]

1 Hochschule für Angewandte Wissenschaften Hamburg
2 Universität Duisburg-Essen
{sebastian.schultheiss, sebastian.suenkler, nurce.yagci, daniela.sygulla, sonja.vonmach, dirk.lewandowski}@haw-hamburg.de

Abstract

Mit dem Result Assessment Tool (RAT) entwickeln wir eine Software zur Durchführung von Studien, bei denen Suchergebnisse kommerzieller Suchmaschinen oder anderer Information-Retrieval-Systeme die Grundlage sind. Ursprünglich entwickelt für Retrievalstudien, bietet das RAT eine Reihe an weiteren Einsatzmöglichkeiten für Forschende aus unterschiedlichen Disziplinen. Dazu zählen die Gesundheitswissenschaften, Medien- und Kommunikationswissenschaften sowie Politikwissenschaften. Basierend auf einer Literaturrecherche stellt dieser Beitrag die Breite der Disziplinen und Studientypen dar, die durch das RAT unterstützt werden und illustriert dies anhand von Fallbeispielen. Durch seine Modularität ist das RAT offen für Bedarfe anderer Forschender, die das RAT selbst nutzen sowie für eigene Zwecke weiterentwickeln möchten.

Keywords: Suchmaschinen; Suchergebnisse; Web Scraping; Retrievaltest; Informationsqualität; Klassifikation; Software; Forschungssoftware

1 Einleitung

Ob in der Informationswissenschaft oder Disziplinen wie Medien- und Kommunikationswissenschaften, Gesundheitswissenschaften oder Politik-

wissenschaften – mit den unterschiedlichsten Zielsetzungen stützen sich Forschende in ihren Studien auf die Ergebnisse kommerzieller Suchmaschinen und anderer Information-Retrieval-Systeme. Neben Retrievalstudien zählen dazu beispielsweise Klassifikationsstudien politischer Informationen oder Qualitätsbewertungen von Patienteninformationen.

Solchen Vorhaben ist gemein, dass mit dem Studiendesign, dem Sammeln und Vorlegen von Suchergebnissen zur Bewertung und schließlich der Auswertung ein erheblicher manueller Aufwand verbunden ist. Vor dem Hintergrund dieser Problematik werden in der Information-Retrieval-(IR-)Community seit Jahren Tools entwickelt. Diese waren jedoch zur einmaligen Nutzung beabsichtigt (z.B. Bar-Ilan/Levene, 2011; Tawileh et al., 2010; Trielli/Diakopoulos, 2022), dienen eng beschränkten Anwendungsfällen (The Digital Methods Initiative, 2020; Thelwall, 2009) oder sind nicht weiterentwickelte Prototypen (Lingnau et al., 2010; Renaud/Azzopardi, 2012) und Software für Testkollektionen (Dussin/Ferro, 2008; Koopman/Zuccon, 2014; Ogilvie/Callan, 2001).

Im Rahmen eines von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts entwickeln wir das *Result Assessment Tool (RAT)*. Das RAT integriert alle Schritte, die bei Studien auf Grundlage von Suchergebnissen zu gehen sind, in einer nachhaltigen Komplettlösung und fördert damit die Reproduzierbarkeit entsprechender Studien.

2 Kurzbeschreibung des RAT

Das RAT ist ein Toolkit, das es Forschenden ermöglicht, Studien auf der Grundlage von Ergebnissen aus Suchmaschinen und anderen Information-Retrieval-Systemen durchzuführen. Zu den aktuell vom RAT unterstützten Suchmaschinen zählen Google, Bing und Ecosia in den Länderversionen Deutschland, Schweden und USA, wobei die Liste der unterstützten Suchmaschinen laufend erweitert wird.

Das RAT wird in Python mit einer PostgreSQL-Datenbank sowie Selenium als Technologie für das Web Scraping entwickelt. Eine detailliertere Beschreibung des technischen Aufbaus des RAT findet sich in Sünkler et al. (2023).

Während Forschende Zugang zu einem Interface haben, um die Studie zu entwerfen, können die Teilnehmenden ebenfalls über ein Web-Interface die

Suchergebnisse zu den vordefinierten Fragen bewerten. Dabei ist das Tool so flexibel gestaltet, dass praktisch alle Studien durchführbar sind, bei denen Suchergebnisse die Grundlage bilden. Neben klassischen Retrievalstudien sind somit Klassifikationsstudien und Datenanalysen oder auch qualitative Inhaltsanalysen möglich. Insgesamt setzt sich das RAT aus fünf Modulen zusammen. Dazu zählen Module zur Testgestaltung und Festlegung der Art der Studie, ein Suchmaschinen-Scraper, der automatisiert Suchergebnisse sammelt, sowie ein Modul zur Fragebogengestaltung. In einem Web-Interface können dann Juroren und Jurorinnen die Fragebögen zu den gesammelten Suchergebnissen ausfüllen. Forschende haben anschließend die Möglichkeit, die Antworten innerhalb eines Analyse-Interfaces auszuwerten und die Ergebnisse herunterzuladen.

3 Einsatzmöglichkeiten des RAT

Die Einsatzmöglichkeiten des RAT wurden zu Projektbeginn anhand einer Literaturrecherche zusammengetragen. Die Recherche dient dem Verständnis und der Beschreibung des Nutzungskontextes (DIN, 2020, S. 21) und stellt somit einen grundlegenden Schritt der Tool-Entwicklung nach den Kriterien der menschzentrierten Gestaltung (User-Centered Design, UCD) dar. In Scopus wurde dazu nach Studien recherchiert, in denen eine Bewertung (*evaluate*, *rate*, *assess* oder andere Synonyme) von Suchergebnissen stattfand. Dies resultierte in Publikationen unterschiedlicher Studien und Disziplinen, wobei $n = 36$ davon näher betrachtet wurden.[1]

Der Großteil der ausgewerteten Studien stammt, wie in Abbildung 1 dargestellt wird, aus den Disziplinen Informationswissenschaft bzw. Information Retrieval (IR), Gesundheits-, Medien- und Kommunikations- sowie Politikwissenschaften.

1 Die Literatursichtung hatte nicht den Anspruch einer systematischen Literaturübersicht, sondern unterschiedliche RAT-Anwendungsmöglichkeiten zusammenzutragen. So bleiben Studien, deren methodisches Vorgehen sich mit bereits recherchierten Arbeiten glich, unberücksichtigt.

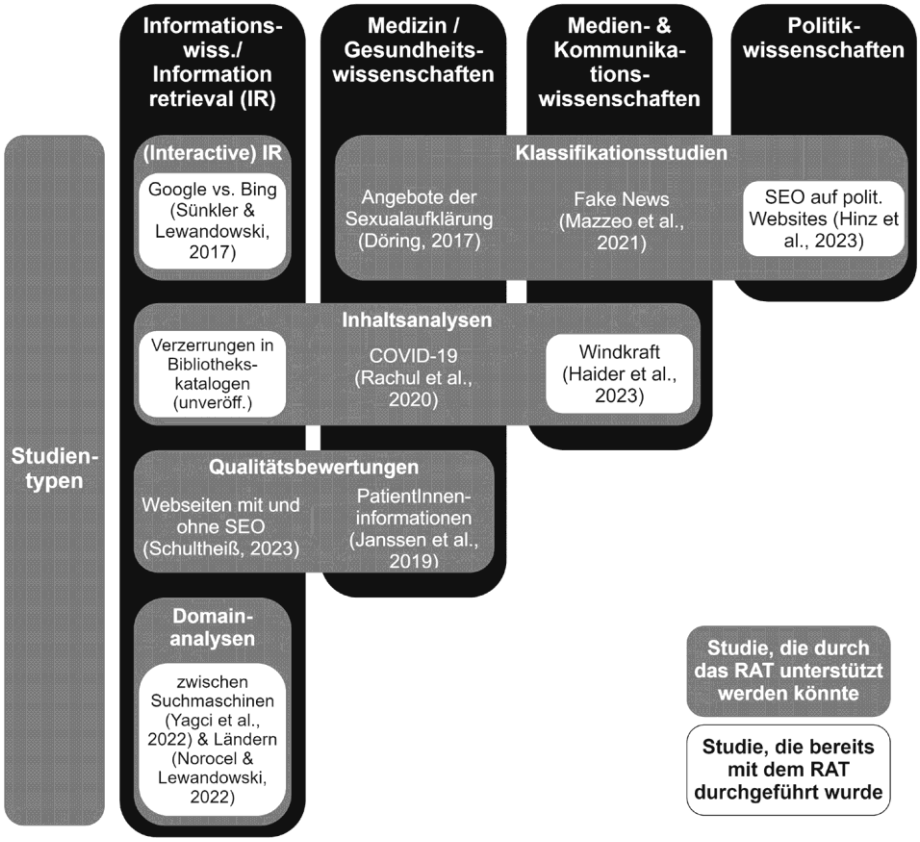

Abb. 1 Studientypen und Disziplinen, die durch das RAT unterstützt werden

Im Folgenden werden die verschiedenen Studientypen anhand von Fallbeispielen illustriert.

3.1 (Interactive) Information Retrieval

Der Anwendungsfall, für den das RAT ursprünglich entwickelt wurde, sind klassische Retrievalstudien wie die Arbeit von Lewandowski (2015) zur Retrieval-Effektivität von Google und Bing. Daneben wird das RAT künftig auch Interactive-Information-Retrieval-(IIR-)Studien ermöglichen. Diese erfolgen unter Einbeziehung von durch ProbandInnen zu lösenden Aufgabenstellungen, während die Interaktionen der Nutzenden erfasst werden (Sünkler/Lewandowski, 2017).

3.2 Klassifikationsstudien

Sowohl manuelle als auch automatisierte Suchergebnisklassifikationen werden durch das RAT unterstützt. So können Forschende anstelle der händischen Sammlung und Bewertung der Suchergebnisse erheblich durch das RAT unterstützt werden (z. B. Döring, 2017; Mazzeo et al., 2021). Mit Blick auf die automatisierte Ergebnisklassifikation ist die Studie von Hinz et al. (2023) zu nennen, welche die RAT-Analysekomponente zur Klassifikation der Suchergebnisse nach Suchmaschinenoptimierungs-(SEO-)Wahrscheinlichkeit einsetzte.

3.3 Inhaltsanalysen

Als weitere RAT-unterstützte Studienform sind Inhaltsanalysen auf Grundlage von Suchergebnissen zu nennen, wie beispielsweise die Arbeit von Rachul et al. (2020) zu COVID-19-bezogenen Suchergebnissen. Hier kann das RAT seine Stärke in der Integration der Studienschritte ausspielen, da der Bedarf weiterer Tools zur inhaltsanalytischen Auswertung entfällt. Zwei Inhaltsanalysen wurden bereits mit dem RAT durchgeführt, in denen Suchergebnisse bezüglich Verzerrungen in Bibliothekskatalogen (studentische Arbeit, unveröffentlicht) bzw. zum Thema „Windkraft“ (Haider et al., 2023) analysiert wurden. In letztgenannter Studie wurde mit der Generierungsmöglichkeit neuer Suchanfragen auf Basis von Ausgangsbegriffen (Schultheiß et al., 2023) eine weitere RAT-Komponente eingesetzt.

3.4 Qualitätsbewertungen

Studien zur Qualitätsbewertung von Online-Informationen sind besonders im Bereich der Patienteninformationen weit verbreitet. Forschende wie Janssen et al. (2019) setzen dazu in der Regel standardisierte Fragebögen wie DISCERN ein. DISCERN besteht aus Fragen wie „Ist die Publikation ausgewogen und unbeeinflusst geschrieben?“, die auf einer fünfstufigen Likert-Skala zu beantworten sind (Charnock et al., 1999) und ohne weiteres im RAT integriert werden können. Neben Gesundheitsinformationen kann die Informationsqualität auch in jedem anderen Forschungskontext untersucht werden, zum Beispiel hinsichtlich der Frage, ob sich Qualitätsunterschiede zwischen Webseiten mit und ohne SEO-Einsatz ermitteln lassen (Schultheiß, 2023).

3.5 Domainanalysen

Domainanalysen sind eine weitere Komponente des RAT. So können Studien zur Analyse von Quellenverteilungen und Domaindichte zwischen Suchmaschinen, Suchanfragen und Ländern durchgeführt werden. Bereits durchgeführte Studien haben Unterschiede zwischen Mainstream- und rechtsradikalen Suchanfragen in einem Ländervergleich zwischen Deutschland und Schweden untersucht (Norocel/Lewandowski, 2023). Eine weitere Studie hat anhand von Google-Trends-Suchanfragen die Unterschiede zwischen Suchmaschinen in Deutschland und den USA analysiert (Yagci et al., 2022).

4 Ausblick

Die oben geschilderten Anwendungsfälle stellen den Status quo des laufenden RAT-Entwicklungsprojekts dar, weshalb kontinuierlich Erweiterungen realisiert werden. So ist eine Erweiterung geplant, die experimentelle Designs im Rahmen von IIR-Studien zulässt, wie beispielsweise eine automatisierte Manipulation der Ergebnisreihenfolge.

Im Sinne der Nachhaltigkeit des RAT-Projekts ist eine starke Einbeziehung der Anwender- und Entwickler-Community unabdingbar. Um dies zu gewährleisten, wird eine AnwenderInnen- und EntwicklerInnen-Community durch verschiedene Maßnahmen wie bspw. eigene AnwenderInnentreffen und Teilnahmen an Konferenzen aufgebaut und gepflegt. Das modulare Design und der öffentlich zugängliche Code des RAT ermöglichen interessierten EntwicklerInnen, weitere Funktionen für das Tool zu programmieren und das Tool ihren Bedürfnissen entsprechend anzupassen.

Danksagung

Die Entwicklung des Result Assessment Tools erfolgt im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts Relevance Assessment Tool (RAT)[2], Förderkennzeichen 460676551.

2 Projekt- und Toolnamen sind nicht mehr identisch, da eine Umbenennung des Tools nach Projektbeginn stattfand.

Literatur

Bar-Ilan, J.; Levene, M. (2011): A method to assess search engine results. In: *Online Information Review*, 35 (6), 854–868. https://doi.org/10.1108/14684521111193166

Charnock, D.; Shepperd, S.; Needham, G.; Gann, R. (1999): DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. In: *Journal of Epidemiology & Community Health*, 53 (2), 105–111. https://doi.org/10.1136/jech.53.2.105

DIN – Deutsches Institut für Normung e. V. (2020): Ergonomie der Mensch-System-Interaktion – Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210:2019). Deutsche Fassung EN ISO 9241-210:2019.

Döring, N. (2017): Sexualaufklärung im Internet. In: *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 60 (9), 1016–1026. https://doi.org/10.1007/s00103-017-2591-0

Dussin, M.; Ferro, N. (2008): Design of a digital library system for large-scale evaluation campaigns. In: B. Christensen-Dalsgaard, D. Castelli, B. Ammitzbøll Jurik, J. Lippincott (Hrsg.): *Research and advanced technology for digital libraries. 12th European Conference. Proceedings / ECDL 2008, Aarhus, Denmark, Sept. 14–19, 2008* (S. 400–401). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-87599-4_45

Haider, J.; Ekström, B.; Tattersall Wallin, E.; Gunnarsson Lorentzen, D.; Rödl, M.; Söderberg, N. (2023): Tracing online information about wind power in Sweden: An exploratory quantitative study of broader trends. https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Ahb%3Adiva-29496

Hinz, K.; Sünkler, S.; Lewandowski, D. (2023): SEO im Wahlkampf: Welche Kandidierende durch Suchmaschinenoptimierung ihre Sichtbarkeit zu erhöhen versuchen. In: K.-R. Korte, M. Schiffers, A. von Schuckmann, S. Plümer (Hrsg.): *Die Bundestagswahl 2021* (S. 1–28). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-35758-0_19-1

Janssen, S.; Fahlbusch, F. B.; Käsmann, L.; Rades, D.; Vordermark, D. (2019): Radiotherapy for prostate cancer: DISCERN quality assessment of patient-oriented websites in 2018. In: *BMC Urology*, 19, Article 42. https://doi.org/10.1186/s12894-019-0474-4

Koopman, B.; Zuccon, G. (2014): Relevation!: An open source system for information retrieval relevance assessment. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, S. 1243–1244. https://doi.org/10.1145/2600428.2611175

Lewandowski, D. (2015): Evaluating the retrieval effectiveness of web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*, 66 (9), 1763–1775. https://doi.org/10.1002/asi.23304

Lingnau, A.; Ruthven, I.; Landoni, M.; Van Der Sluis, F. (2010): Interactive search interfaces for young children—The PuppyIR approach. In: *Proceedings – 10th IEEE International Conference on Advanced Learning Technologies, ICALT 2010*, S. 389–390. https://doi.org/10.1109/ICALT.2010.111

Mazzeo, V.; Rapisarda, A.; Giuffrida, G. (2021): Detection of Fake News on COVID-19 on Web Search Engines. In: *Frontiers in Physics*, 9, Article 685730. https://doi.org/10.3389/fphy.2021.685730

Norocel, O. C.; Lewandowski, D. (2023): Google, data voids, and the dynamics of the politics of exclusion. *Big Data & Society*, 10 (1). https://doi.org/10.1177/20539517221149099

Ogilvie, P.; Callan, J. P. (2001): Experiments Using the Lemur Toolkit. In E. M. Voorhees, D. K. Harman (Hrsg.): *Proceedings of The Tenth Text REtrieval Conference, TREC 2001, Gaithersburg, Maryland, USA*. National Institute of Standards and Technology (NIST).

Rachul, C.; Marcon, A. R.; Collins, B.; Caulfield, T. (2020): COVID-19 and 'immune boosting' on the internet: A content analysis of Google search results. In: *BMJ Open*, 10, Article e040989. https://doi.org/10.1136/bmjopen-2020-040989

Renaud, G.; Azzopardi, L. (2012): SCAMP: A tool for conducting interactive information retrieval experiments. In: *IIiX 2012 – Proceedings 4th Information Interaction in Context Symposium: Behaviors, Interactions, Interfaces, Systems*, S. 286–289. https://doi.org/10.1145/2362724.2362776

Schultheiß, S. (2023): How search engine marketing influences user knowledge gain: Development and empirical testing of an information search behavior model. In: *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23), March 19–23, 2023, Austin, TX, USA*. https://doi.org/10.1145/3576840.3578297

Schultheiß, S.; Lewandowski, D.; Mach, S. von; Yagci, N. (2023): Query sampler: Generating query sets for analyzing search engines using keyword research tools. In: *PeerJ Computer Science*, 9, Article e1421. https://doi.org/10.7717/peerj-cs.1421

Sünkler, S.; Lewandowski, D. (2017): Does it matter which search engine is used? A user study using post-task relevance judgments. In: *Proceedings of the Association for Information Science and Technology*, 54 (1), 405–414. https://doi.org/10.1002/pra2.2017.14505401044

Sünkler, S.; Yagci, N.; Sygulla, D.; Mach, S. von; Schultheiß, S.; Lewandowski, D. (2023): Result Assessment Tool: Software-Toolkit für die Durchführung von Studien auf der Grundlage von Suchergebnissen. In: W. Semar (Hrsg.): *Nachhaltige*

Information – Information für Nachhaltigkeit. Tagungsband des 17. Internationalen Symposiums für Informationswissenschaft, Chur, Schweiz 7. bis 9. Nov. 2023 (S. 438–444). Glückstadt: Verlag Werner Hülsbusch. https://dx.doi.org/10.5281/zenodo.10009338

Tawileh, W.; Mandl, T.; Griesbaum, J. (2010): Evaluation of five web search engines in arabic language. In: M. Atzmüller, D. Benz, A. Hotho, G. Stumme (Hrsg.): *LWA 2010– Lernen, Wissen und Adaptivitat – Learning, Knowledge, and Adaptivity, Workshop Proceedings* (S. 221–228).

The Digital Methods Initiative (2020): *ToolDatabase < Dmi < Foswiki*. https://wiki.digitalmethods.net/Dmi/ToolDatabase

Thelwall, M. (2009): *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. Cham: Springer International Publishing. https://link.springer.com/book/10.1007/978-3-031-02261-6

Trielli, D.; Diakopoulos, N. (2022): Partisan search behavior and Google results in the 2018 U.S. midterm elections. In: *Information, Communication & Society*, 25 (1), 145–161. https://doi.org/10.1080/1369118X.2020.1764605

Yagci, N.; Sünkler, S.; Häußler, H.; Lewandowski, D. (2022): A Comparison of Source Distribution and Result Overlap in Web Search Engines. In: *Proceedings of the 85th Annual Meeting of the Association of Information Science and Technology, Pittsburgh, PA*. https://doi.org/10.1002/pra2.758

In: W. Semar (Hrsg.): Nachhaltige Information – Information für Nachhaltigkeit. Tagungsband des 17. Internationalen Symposiums für Informationswissenschaft (ISI 2023), Chur, Schweiz, 7.–9. November 2023. Glückstadt: Verlag Werner Hülsbusch, S. 429–437. DOI: https://dx.doi.org/10.5281/zenodo.10009338