

Result Assessment Tool

Software-Toolkit für die Durchführung von Studien auf der Grundlage von Suchergebnissen

Sebastian Sünkler¹, Nurce Yagci¹, Daniela Sygulla¹, Sonja von Mach¹, Sebastian Schultheiß¹, Dirk Lewandowski^{1,2}

¹Hochschule für Angewandte Wissenschaften Hamburg, ²Universität Duisburg-Essen

Einleitung

Die Durchführung umfangreicher und reproduzierbarer Studien, die sich auf die Ergebnisse kommerzieller Suchmaschinen und anderer Information-Retrieval-(IR-)Systeme stützen, ist aufgrund verschiedener Aspekte schwierig. Zu diesen Aspekten zählen insbesondere die **Sammlung von Suchergebnissen**, die **Suche nach Jurorinnen und Juroren** sowie die **Erfassung ihrer Bewertungen zu den Suchergebnissen** und die anschließenden **Datenanalysen**. Dies erfordert in der Regel zeitintensive und aufwendige manuelle Maßnahmen und Schritte. Um diese Probleme anzugehen und um Studien mit Suchergebnissen durchzuführen, werden in der IR-Community seit Jahren Software-Tools entwickelt. Dabei handelt es sich jedoch in erster Linie um Tools für den einmaligen Gebrauch ([1]-[3]), um Prototypen, die nicht weiterentwickelt wurden ([4], [5]), und um Software für Testsammlungen ([6], [7]) oder eng beschränkte Anwendungsfälle ([8], [9]). Zusammenfassend fehlt es bisher an einer nachhaltigen Software-Lösung, um alle Schritte für solche Studien zu integrieren. Mit dem **Result Assessment Tool (RAT) entwickeln wir ein Software-Toolkit**, das die genannten Schritte für **die Durchführung von groß angelegten Studien mit Suchergebnissen vereinfachen soll**. In dem folgenden Abschnitt stellen wir die Funktionsweise aller bisherigen Komponenten vor.

Funktionalität

Das **RAT ist eine flexible webbasierte Software**, die in **Python** mit einer **PostgreSQL-Datenbank** sowie **Selenium als Technologie für das Web Scraping** entwickelt wird. Die Architektur der Software teilt sich dabei in Web Frontend, das in Flask entwickelt wird, und ein Server Backend auf, in dem alle Prozesse zur Datenerfassung (vor allem das Scraping der Suchmaschinen), Prozesse für automatisierte Analysen (z. B. Klassifikationen von Suchergebnissen) sowie die Datenbankoperationen durchgeführt werden. Bei dem Web Frontend stehen zwei Anwendungen zur Verfügung. In der einen Anwendung haben Forschende Zugang zu einem Interface, um Studien zu entwerfen, während die andere Anwendung den Zugang für Juroren und Jurorinnen bereitstellt, in dem Suchergebnisse bewertet werden. 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 Result Assessment Tool aus sechs Modulen zusammen, die die Durchführung der Studien ermöglichen:

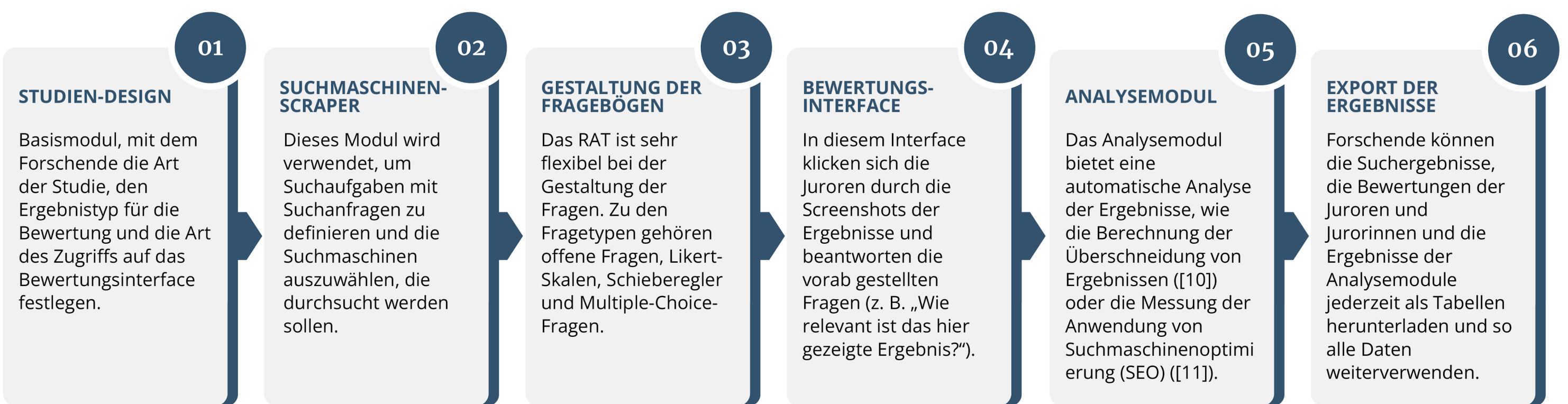


Abb 1. RAT-Module

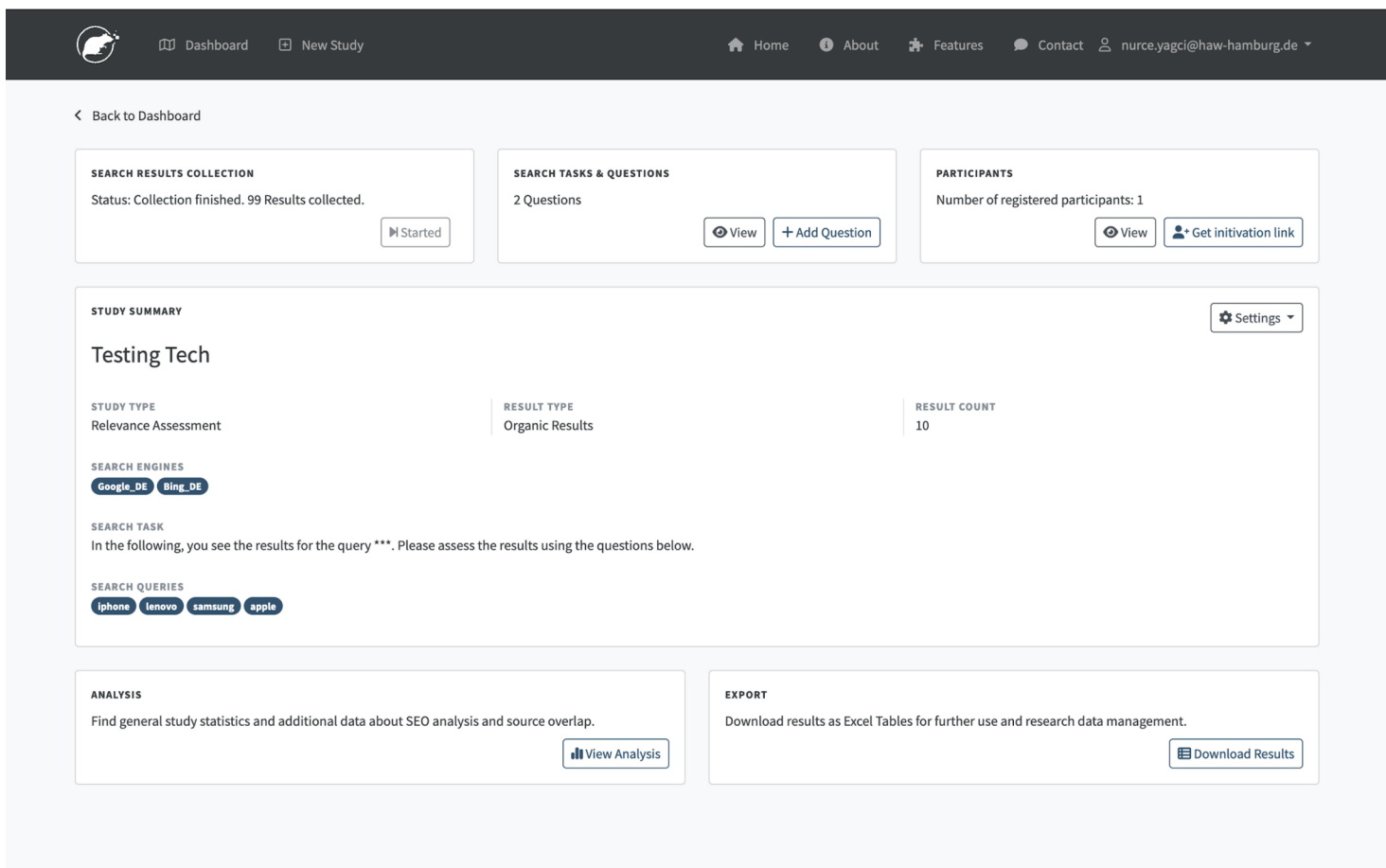


Abb 2. RAT-Dashboard

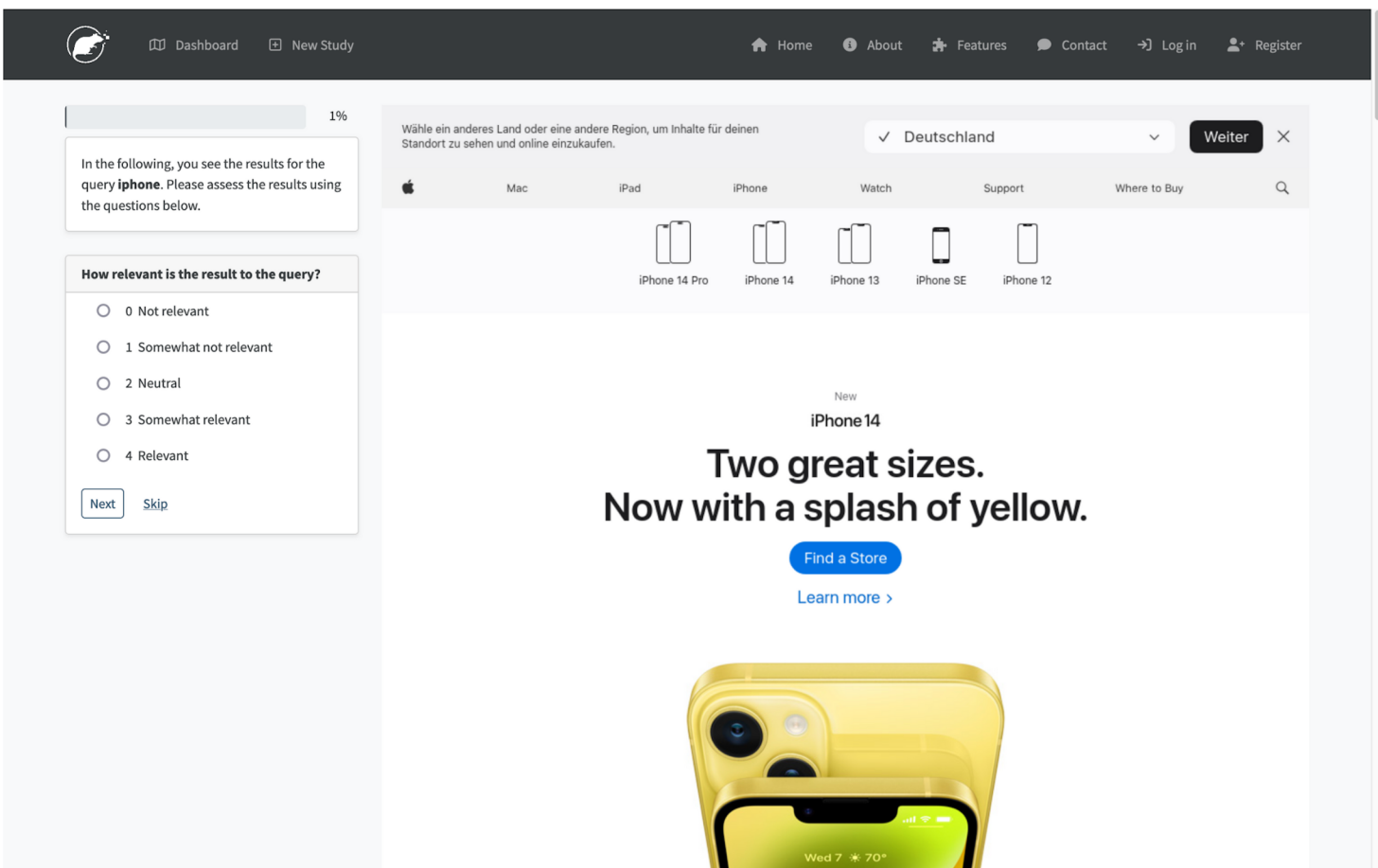


Abb 3. RAT-Bewertungsinterface

Abb. 2 zeigt die Übersichtsseite in der Anwendung für die Forschenden im Web Frontend. Diese Ansicht zeigt den **aktuellen Status der Studie und bietet eine Übersicht zu den Suchmaschinen, Suchanfragen, Aufgabenstellungen und der Anzahl der bisherigen Teilnehmenden**. Die Studienzusammenfassung gibt einen Überblick über alle Optionen, die im Testdesignprozess festgelegt wurden, wie die Suchanfragen und Suchmaschinen. Abb. 3 zeigt ein **Beispiel für eine Bewertung in der Bewertungsoberfläche des Result Assessment Tool**. Die Juroren und Jurorinnen sehen auf der linken Seite unter der jeweiligen Suchanfrage alle festgelegten Fragen und auf der rechten Seite einen Screenshot eines zu bewertenden Ergebnisses.

Alle gesammelten Ergebnisse und Bewertungen lassen sich durch automatische Analyseprozesse im RAT weiterverarbeiten. Dabei werden Statistiken über die Studie berechnet und berichtet. Diese Statistiken geben einen Überblick über die Anzahl der Suchanfragen, die zu sammelnden Suchergebnisse pro Anfrage und die erwarteten und gesammelten Ergebnisse, die Suchmaschinenbedingt von der gewünschten Suchergebnismenge abweichen kann. Zusätzlich lassen sich **weitere optionale automatisierte Analysen durchführen**, z. B. die Berechnung der **Wahrscheinlichkeit der Verwendung von SEO für ein Suchergebnis** oder die Berechnung der **Überschneidung der Suchergebnisse zwischen den in einer Studie verwendeten Suchmaschinen**. Diese beiden Beispiele zeigen bereits das Potenzial der automatisierten Analysen.

Evaluierung des Result Assessment Tools

Die Entwicklung der Software erfolgt gemäß **den Standards der menschenzentrierten Gestaltung** (User-Centered Design; UCD) ([12]). Die Bewertung der Benutzerfreundlichkeit (Usability) der Software ist ein wesentlicher Bestandteil des UCD-Prozesses. Aus diesem Grund werden **kontinuierliche Usability-Tests und heuristische Evaluationen** verwendet.

Verfügbarkeit der Software-Demo, Quellcode und Forschungsdaten

Eine Demo der RAT kann von Forschenden unter <https://rat-software.org/> genutzt werden. Wir folgen bei der Veröffentlichung der Forschungsdaten, zu Studien, die wir selbst mit dem RAT durchführen und bei der Veröffentlichung des Quellcodes der Software den FAIR-Prinzipien ([13]). Die Forschungsdaten können unter <https://osf.io/t3hg9/> und der Quelltext unter <https://github.com/rat-software/> eingesehen werden.

Danksagung

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

Referenzen

- [1] Bar-Ilan, J., & Levene, M. (2011). A method to assess search engine results. Online Information Review, 35(6), 854–868. <https://doi.org/10.1108/1468452111193166>.
- [2] Tawileh, W., Griesbaum, J., & Mandl, T. (2010). Evaluation of five web search engines in Arabic language. In M. Atzmüller, D. Benz, A. Hotho, & G. Stumme (Eds.), Proceedings of LWA2010 (pp. 1–8).
- [3] Trielli, D., & Diakopoulos, N. (2020). Partisan search behavior and Google results in the 2018 U.S. midterm elections. Information, Communication & Society, 0(0), 1–17. <https://doi.org/10.1080/1369118X.2020.1764605>.
- [4] Lingnau, A., Ruthven, I., Landoni, M., & van der Sluis, F. (2010). Interactive Search Interfaces for Young Children - The PuppyIR Approach. 2010 10th IEEE International Conference on Advanced Learning Technologies, 389–390. <https://doi.org/10.1109/ICALT.2010.111>.
- [5] Renaud, G., & Azzopardi, L. (2012). SCAMP. Proceedings of the 4th Information Interaction in Context Symposium on - IIIX '12, 286–289. <https://doi.org/10.1145/2362724.2362776>.
- [6] Dussin, M., & Ferro, N. (2008). Design of a Digital Library System for Large-Scale Evaluation Campaigns. In B. Christensen-Dalsgaard, D. Castellì, B. Ammitzbøll Jurik, & J. Lippincott (Eds.), Research and Advanced Technology for Digital Libraries (pp. 400–401). Springer Berlin Heidelberg.
- [7] Koopman, B. (2014). Semantic Search as Inference. ACM SIGIR Forum. <https://doi.org/10.1145/2701583.2701601>.
- [8] The Digital Methods Initiative. (2022). DMI Tools. <https://wiki.digitalmethods.net/Dmi/ToolDatabase>.
- [9] Thelwall, M. (2009). Introduction to Webometrics: Quantitative Web Research for the Social Sciences. Synthesis Lectures on Information Concepts, Retrieval, and Services. <https://doi.org/10.2200/s00176ed1v01y200903icr004>.
- [10] Yagci, N., Sünkler, S., Häußler, H., & Lewandowski, D. (2022). A Comparison of Source Distribution and Result Overlap in Web Search Engines. Proceedings of the Association for Information Science and Technology, 59(1), 346–357. <https://doi.org/10.1002/pra2.758>.
- [11] Lewandowski, D., Sünkler, S., & Yagci, N. (2021). The influence of search engine optimization on Google's results: A multi-dimensional approach for detecting SEO. 13th ACM Web Science Conference 2021 (WebSci '21), June 21–25, 2021, Virtual Event, United Kingdom. <https://doi.org/10.1145/3447535.3462479>.
- [12] Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. In W. Bainbridge (Hrsg.), Encyclopedia of Human-Computer Interaction (S. 445–456). Sage Publications.
- [13] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.



SEARCH
STUDIES

HAW
HAMBURG

