

FUNDED BY

DFG Deutsche
Forschungsgemeinschaft

HANDS ON RAT

ANALYSE VON SUCHSYSTEMEN MIT DEM SOFTWARETOOL



112.BIBLIO CON 20
24
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

AGENDA

1. Einführung
2. Das RAT-Projekt
3. Was ist das RAT und wie kann es genutzt werden?
4. Studien mit dem RAT
5. Gemeinsame Studie
6. Abschluss



1. EINFÜHRUNG



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

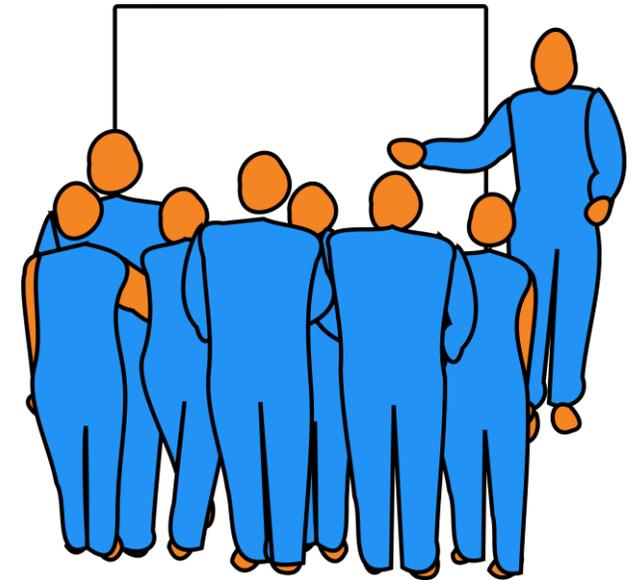
1. EINFÜHRUNG

Was werden Sie in diesem Tutorial lernen?

Sie werden erfahren, wie das RAT **funktioniert** und in der Praxis **eingesetzt** werden kann.

Was werden wir heute zusammen erleben?

Wir werden eine Studie in RAT **entwerfen** und **durchführen**.



2. DAS RAT-PROJEKT



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

2. DAS RAT-PROJEKT

Das Projektziel

- Forschenden mit Hilfe einer Software den **Zugriff** auf und die Analyse von **Suchmaschinendaten** (= Ergebnisse von verschiedenen Suchmaschinen) zu ermöglichen

Motivation für das Projekt

- **Kein offizieller Zugang** zu Suchmaschinendaten
- Die einzige verfügbare API ist die von **Bing**
- Daten für Studien werden oft **manuell gesammelt**, was zu kleinen Studien führt
- Software für die Sammlung der Suchergebnisse wird oft nur für eine einzige Studie entwickelt und ist schnell **veraltet und nicht flexibel**



2. DAS RAT-PROJEKT

Forschungsprojekt durchgeführt durch die Forschungsgruppe Search Studies (searchstudies.org)

Gefördert von der Deutschen Forschungsgemeinschaft (DFG), 2021-2024.

Projektbestandteile

- **Entwicklung** von Forschungs-Software
- **Unterstützung** für interessierte Forschende
- **Community building**: Förderung der Suchmaschinenforschung



3. WAS IST DAS RAT UND WIE KANN ES GENUTZT WERDEN?



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

3. WAS IST DAS RAT?

Module im RAT

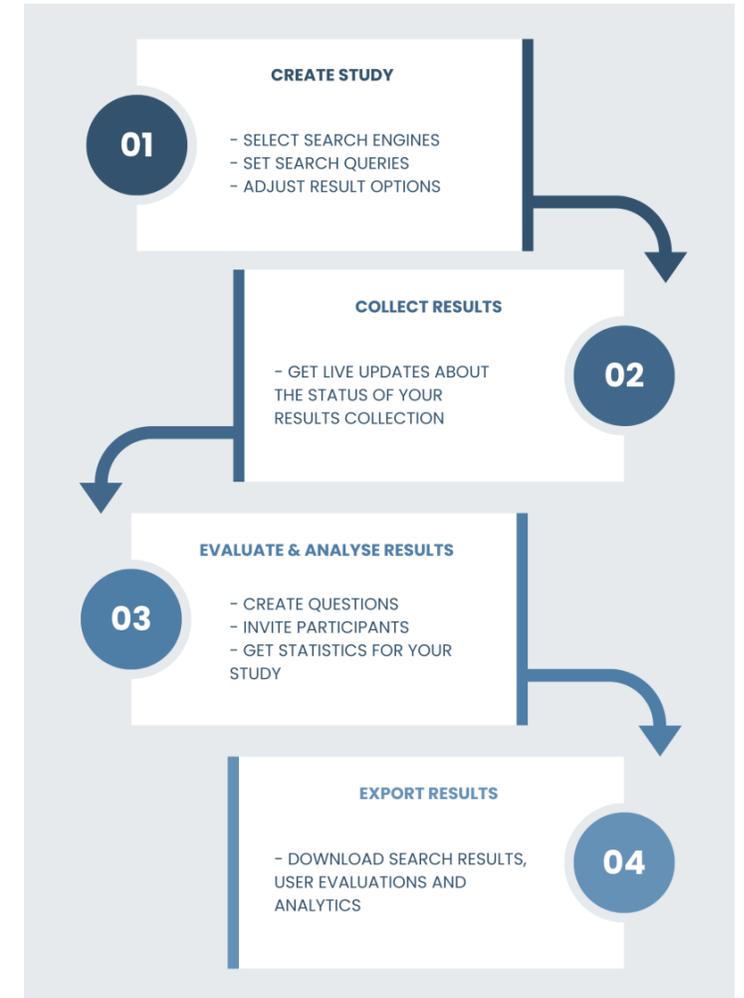
- Studiendesign, Suchmaschinen-scraping, jurorenbasierte Evaluation, automatische Ergebnisanalyse, Download der Ergebnisse

Scraper

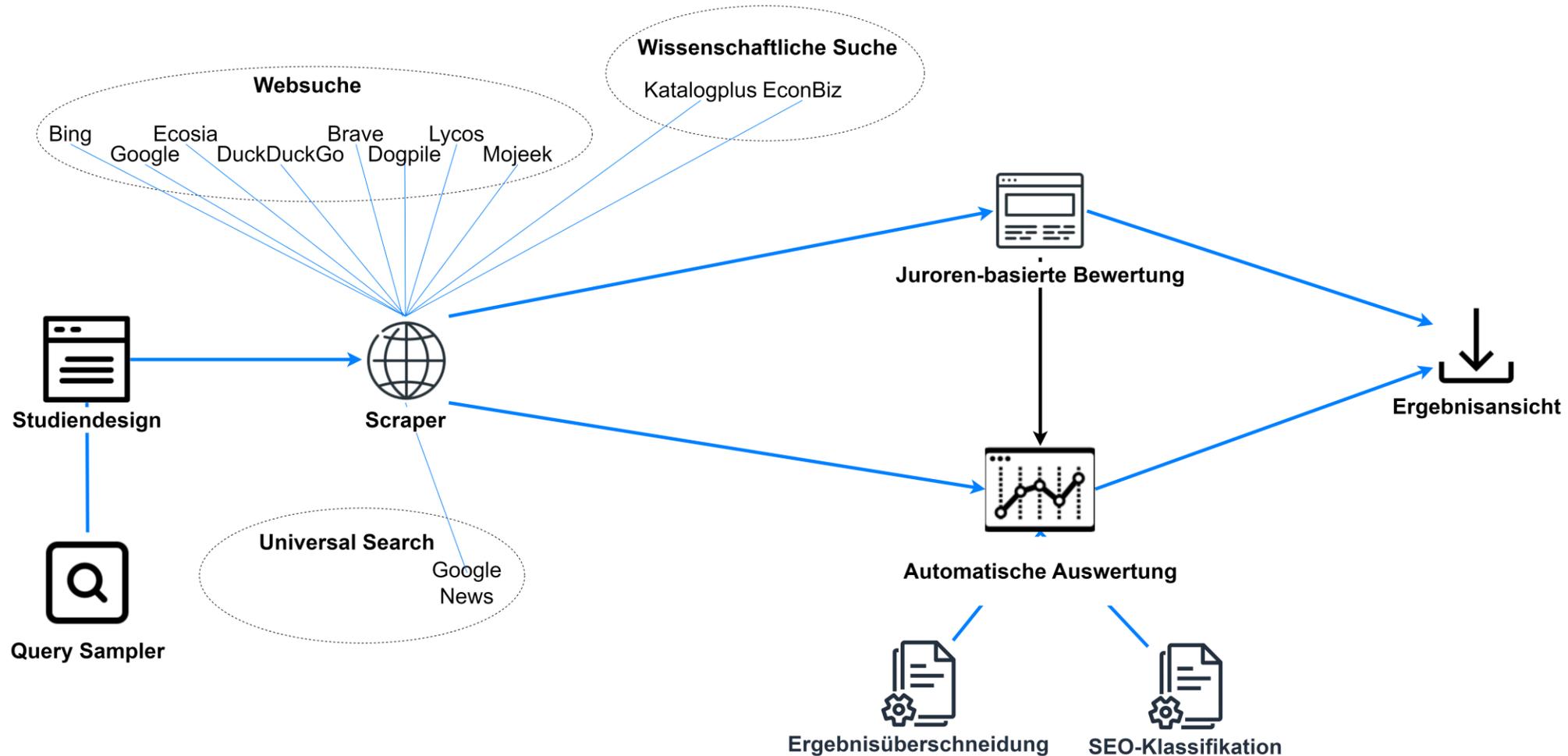
- Automatisches Speichern der Ergebnisse von Suchsystemen (z. B. Google, Bibliothekskataloge, E-Commerce-Portale)
- RAT kann potenziell jede Webseite verbunden werden, die ein Suchfeld hat.

Erweiterungen

- Analysekomponenten und Scraper können durch das modulare Design hinzugefügt werden



3. WAS IST DAS RAT? MODULE & SUCHMASCHINEN



3. WAS IST DAS RAT? DASHBOARD FÜR FORSCHENDE

SEARCH RESULTS COLLECTION
Status: Collection finished. 16 Results collected.
Scraper ▶ Started

SEARCH TASKS & QUESTIONS
1 Question
Fragebogen 👁 View + Add Question

PARTICIPANTS
Number of registered participants: 4
Juroren 👁 View + Get invitation link

STUDY SUMMARY ⚙ Settings

Gesundheitsinformationen

STUDY TYPE
Relevance Assessment

RESULT TYPE
Organic Results

RESULT COUNT
3

SEARCH ENGINES
Google_DE Bing_DE Google_US

SEARCH TASK
Sie sehen im Folgenden Ergebnisse für die Suchanfrage ***

SEARCH QUERIES
herzinfarkt covid depression

ANALYSIS
Find general study statistics and additional data about SEO analysis and source overlap.
Statistiken und Analyseergebnisse 📊 View Analysis

EXPORT
Download results as Excel Tables for further use and research data management.
Download der Ergebnisse 📄 Download Results

Zusammenfassung der Suchaufgaben und Suchanfragen

10



112.BIBLIO CON 2024
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

SEARCH STUDIES

HAW HAMBURG

3. WAS IST DAS RAT? BEWERTUNGSINTERFACE

The screenshot displays the RAT interface. At the top, there is a navigation bar with links for Dashboard, New Study, Home, About, Features, Contact, Log in, and Register. Below this, a progress bar shows 90% completion. The main content area is divided into two columns. The left column contains a text box with the instruction: "In the following, you see the results for the query **bim haw**. Please assess the results using the questions below." followed by a blue button labeled "Aufgabe". Below this is a question: "Ist das Ergebnis relevant für die Suchanfrage?" with two radio button options: "nein" and "ja", and a blue button labeled "Fragen". At the bottom of this column are "Next" and "Skip" buttons. The right column features a large blue header for "HAW HAMBURG" with navigation links for "HOCHSCHULE", "STUDIUM", "FORSCHUNG", and "INTERNATIONAL". Below the header is a large image of two students, one holding a camera. Overlaid on the image is the text "BIBLIOTHEKS- UND INFORMATIONSMANAGEMENT" and "Bachelor of Arts (B.A.)". At the bottom of the image are two buttons: "Studieninteressierte" and "Studierende". Below the image is a breadcrumb trail: "STARTSEITE — STUDIUM — BIBLIOTHEKS- UND INFORMATIONSMANAGEMENT" followed by a blue button labeled "Screenshot des Ergebnisses" and a "ZURÜCK" button.



3. WAS IST DAS RAT? AUTOMATISCHE ANALYSE

< Back to Study

RESULTS STATISTICS

Queries	3
Search Engines	3
Results to Collect per Query	3
Collection Status	Zusammenfassung des Scrapings 67 %

EVALUATION STATISTICS

Participants	4
Questions	1
Evaluations per Participant	18
Evaluation Status	Zusammenfassung der Bewertungen 11 %

SEO ANALYSIS ⓘ

Class	Total	Percentage
Most probably optimized	15	0.83
Probably optimized	0	0.0
Probably not optimized	0	0.0
Most probably not optimized	1	0.06
Error	SEO-Analyse 2	0.11

RESULTS OVERLAP



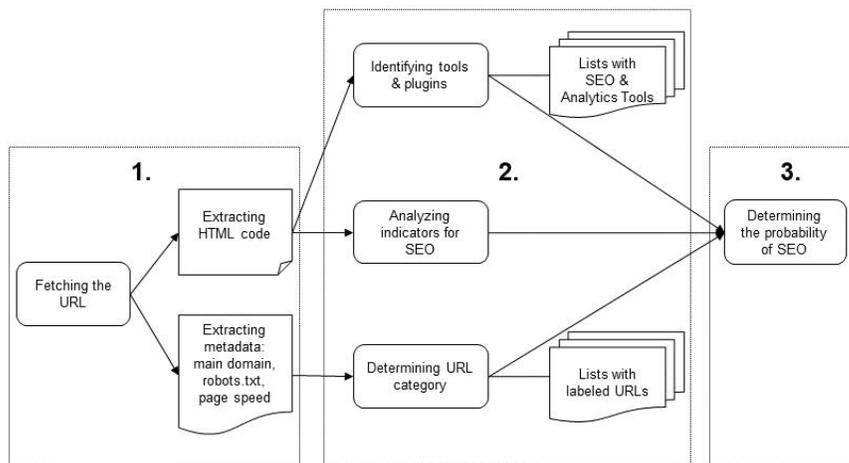
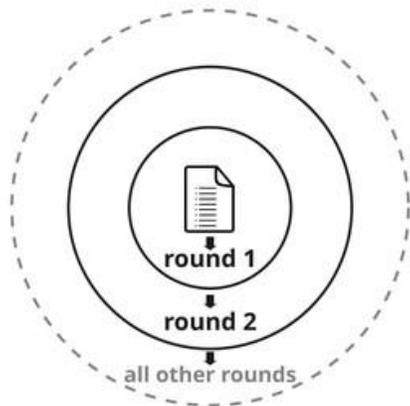
Search Engine Pair	SE 1 exclusive	SE 2 exclusive	Overlap	Total
Google_DE-Bing_DE	0	9	0	9
Google_DE-Google_US	0	9	0	9
Bing_DE-Google_US	9	9	0	18

Ergebnisüberschneidungen



3. WAS IST DAS RAT? ZUSATZMODULE

(D) Generating keyword ideas



Query Sampler

Der Query Sampler ist eine von uns entwickelte Software, um Suchanfragen von kommerziellen Suchmaschinen mit Hilfe von Keyword-Recherche-Tools zu untersuchen.

→ Automatisches Erstellen von **Suchanfragensets**

SEO Classifier

Mit dem SEO Classifier kann der Grad der **Suchmaschinenoptimierung** auf einer URL ermittelt werden.

3. WAS IST DAS RAT? QUERY SAMPLER

Study: Searchstudies

INPUT KEYWORD LIST

Keyword	Language	Region
search studies haw	German	Germany
searchstudies	German	Germany

Number of generated Keyword ideas: 2

[Download Keyword Ideas](#)

INSERT KEYWORDS

LANGUAGE
German

REGION
Germany

[Submit](#)

Copyright © 2024 Result Assessment Tool Developed by Search Studies (HAW Hamburg) Funded by the German Research Foundation (DFG)



4. STUDIEN MIT DEM RAT



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

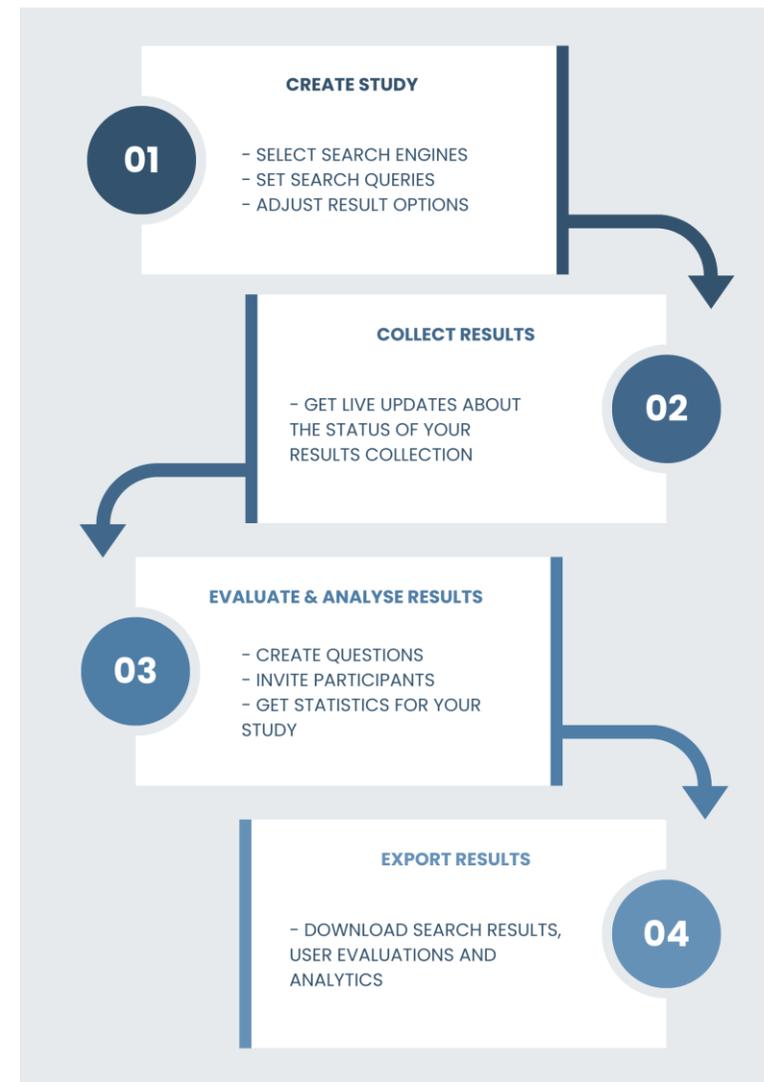
**SEARCH
STUDIES**



**HAW
HAMBURG**

4. STUDIEN MIT DEM RAT: ABLAUF IM RAT

Studien in RAT durchlaufen vier Schritte, darunter das **Studiendesign**, die **Sammlung** von **Suchergebnissen**, die **Bewertung** durch die Teilnehmer und die **Analyse** der Suchergebnisse und Studienergebnisse.



4. STUDIEN MIT DEM RAT: MOMENTANE FEATURES

- Scraping von **Google, Microsoft Bing, Ecosia** und **Katalogplus**
- Regelbasierte **Klassifikation von SEO**
- Definition von **Fragen mit einer Auswahl von Antworttypen:**
 - Offene Fragen
 - Bewertungsskalen
 - Einfach- und Mehrfachauswahl
 - Schieberegler
- Sammlung der Antworten in einem **Bewertungsinterface**
- **Statistiken** zu den Suchergebnissen
- **Herunterladen von Ergebnissen** für weitere Auswertungen



4. STUDIEN MIT DEM RAT: STUDIENTYPEN

- **Retrievalstudien:** Qualität der Suchergebnisse von verschiedenen Suchdiensten
- **Inhaltsanalysen:** Inhalte der Suchergebnisse und Suchergebnisseiten
- **Trefferanalysen:** Überschneidungen von Suchergebnissen, Suchergebnisverteilungen, Suchmaschinenoptimierung



4. STUDIEN MIT DEM RAT: BEISPIELE

Information Retrieval

Evaluating the Retrieval Effectiveness of Web Search Engines Using a Representative Query Sample

Dirk Lewandowski
Department of Information, Hamburg University of Applied Sciences, Finkenau 35, Hamburg D-22081, Germany. E-mail: dirk.lewandowski@haw-hamburg.de

Search engine retrieval effectiveness studies are usually small scale, using only limited query samples. Furthermore, queries are selected by the researchers. We address these issues by taking a random representative sample of 1,000 informational and 1,000 navigational queries from a major German search engine and comparing Google's and Bing's results based on this sample. Juron were found through crowdsourcing, and data were collected using specialized software, the **Relevance Assessment Tool (RAT)**. We found that although Google outperforms Bing in both query types, the difference in the performance for informational queries was rather low. However, for navigational queries, Google found the correct answer in 85.2% of cases, whereas Bing only found the correct answer 76.6% of the time. We conclude that search engine performance on navigational queries is of great importance, because here in this case can clearly identify queries that have returned correct results. So, performance on this query type may contribute to explaining user satisfaction with search engines.

Introduction

Given the importance of search engines in daily use (Purcell, Brenner, & Raine, 2012; Van Eimeren & Fries, 2012), which is also expressed by the many millions of queries entered into general-purpose web search engines every day (comScore, 2010), it is surprising to find how low the number of research articles on search engine quality still is. One might assume that large-scale initiatives to monitor the quality of results from general-purpose search engines, such as Google and Bing, already exist. Users could refer to such studies to decide which search engine is most capable of retrieving the information they are looking for. Yet, as we will demonstrate, existing search engine retrieval

effectiveness studies have covered a much smaller sample size far. This small size has cast a degree of doubt on their conclusions. One could argue that it is search engine vendors' sole responsibility to conduct tests on the retrieval effectiveness of their engines in relation to their competitors (and they surely conduct such studies). The aim of studies conducted by the vendors and academic studies, however, are quite different. Whereas the aim of studies conducted by search engine vendors is to improve a certain search engine (and test additional capabilities, such as dealing with search engine spam), the purpose of academic studies is to advance retrieval effectiveness testing and provide a valid picture of the quality of different search engines. Decision making, reviewing, for instance, around funding alternative search engines, is facilitated with this information. Although search engines are using click-through data, to a large degree, to improve their services and existing research focuses on this area (Chaklin & Sedyukov, 2013; Joachims, Granku, Pan, Hembrooke, & Gay, 2005; Joachims et al., 2007), search engine providers still conduct classic retrieval effectiveness tests using human jurors recruited for that task (Google, 2012). Retrieval effectiveness tests are just one part of the overall evaluation of search engine quality (Lewandowski & Hochstetler, 2008). Measuring the effectiveness of search engine results is an integral part of such frameworks and is usually based on a Text RElevance conference (TREC)-style setting. However, these methods have been criticized as being inadequate for evaluating web search engines, where results presentation differs from other information retrieval systems (Lewandowski, 2013a). Furthermore, retrieval effectiveness tests have been criticized for not taking into account actual user behavior, that is, they are based on the query-response paradigm and do not consider interaction in the searching process (Ingwersen & Järvelin, 2005). Though these criticisms should not be ignored, continued enhancement of query-response-based methods is still a valuable pursuit.

Received August 14, 2019; revised March 19, 2024; accepted March 19, 2024.
© 2015 ASSIR & Published online 23 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23304

JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 66(9):1763–1775, 2015

<https://arxiv.org/abs/1405.2210>

Qualität von Bibliothekssystemen

A framework for designing retrieval effectiveness studies of library information systems using human relevance assessments

Christiane Behrnt, Dirk Lewandowski

Journal of Documentation
ISSN: 0022-0418

Article publication date: 8 May 2017

DOWNLOADS
2089

Abstract

Purpose

The purpose of this paper is to demonstrate how to apply traditional information retrieval (IR) evaluation methods based on standards from the Text REtrieval Conference and web search evaluation to all types of modern library information systems (LISs) including online public access catalogues, discovery systems, and digital libraries that provide web search features to gather information from heterogeneous sources.

Design/methodology/approach

The authors apply conventional procedures from IR evaluation to the US context considering the specific characteristics of modern library materials.

Findings

The authors introduce a framework consisting of five parts: search queries, search results, assessors, testing, and data analysis. The authors show how to deal with comparability problems resulting from diverse document types, e.g. electronic articles vs printed monographs and what issues need to be considered for retrieval tests in the library context.

Practical implications

The framework can be used as a guideline for conducting retrieval effectiveness studies in the library context.

Originality/value

Although a considerable amount of research has been done on IR evaluation and standards for conducting retrieval effectiveness studies do exist, to the authors' knowledge this is the first attempt to provide a systematic framework for evaluating the retrieval effectiveness of contemporary LISs. The authors demonstrate which issues must be considered and what

Related articles

A library information retrieval effectiveness case study
Robert Marjan et al., Libr

Comparative evaluation of health information retrieval
Carla Teixeira Lopes et al.,

Treasure on the Web: current and trends
Ali Akbar Shir et al., Onli

Study on using deep learning to retrieve the biomass of sea from GEO data
Sun Danyang et al., And 2

What will you see when you de/federated? (poster)
Sponsored by Boehinger

Dynamic evaluation of the information food in Xinjiang
Zhou Gang et al., And 2

Powered by TREND|MD

<https://www.emerald.com/insight/content/doi/10.1108/JD-08-2016-0099/full/html>

Verzerrungen in Bibliothekssystemen

Original Research Article

Google, data voids, and the dynamics of the politics of exclusion

Ov Cristian Noroel¹ and Dirk Lewandowski^{2,3}

Abstract

This study deploys a critical approach to big data analytics to gauge the tentative contours of data voids in Google searches that reflect extreme-right dynamics of exclusion in the aftermath of the 2015 humanitarian crisis in Europe. The study adds complexity to the analysis of data voids, expanding the framework of investigation outside the USA context by concentrating on Germany and Sweden. Building on previous big data analytics addressing the politics of exclusion, the study proposes a catalogue of queries concerning the issue of migration in both Germany and Sweden on a continuum from mainstream to extreme-right vocabularies. The catalogue of queries enables specific and localized queries to identify data voids. The results show that a search engine's reliance on source popularity may lead to extreme-right sources appearing in top positions. Furthermore, using platforms for user-generated content provides a way for localized queries to gain top positions.

Keywords

Critical data studies, data void, Germany, politics of exclusion, search engines, Sweden

This article is part of special theme on The State of Google Critique and Intervention. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/psj/psjby/collections/stateofgooglecritiqueandintervention>

Introduction

Among general-purpose web search engines, Google is the most popular in the Western hemisphere, with a total of approximately 80.8 billion visits in early 2022 (SimilarWeb, 2022). This ubiquity of Google search and its seeming ease of use in people's daily lives (Haider and Sundin, 2019; Sundin et al., 2017) obfuscates the low level of search engine knowledge among most ordinary users, which opens the door for the manipulation of propaganda and disinformation. One important vulnerability is described in research as a data void (Goležalov and Boyd, 2019). A data void becomes possible when, in response to a specific query, search engines return skewed and manipulatory content, which, from the point of view of social relevance (Haider and Sundin, 2019), is of low quality because there is hardly any high-quality content corresponding to that search. Of interest here are two types of data voids. First are problematic queries, which concern

¹Department of Gender Studies, Lund University, Lund, Sweden
²Department of Information, Hamburg University of Applied Sciences, Hamburg, Germany
³Department of Computer Science and Applied Cognitive Science, University Duisburg-Essen, Duisburg, Germany

Corresponding author:
Ov Cristian Noroel, Department of Gender Studies, Lund University, Box 117, SE-221 00 Lund, Sweden.
Email: ov.cristian.noroel@gu.se

© 2015 Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>), which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://sagepub.com/journalsPermissions.nav>).

<https://journals.sagepub.com/doi/full/10.1177/20539517221149099>

Politische Kommunikation

SEO im Wahlkampf

Welche Kandidierende durch Suchmaschinenoptimierung ihre Sichtbarkeit zu erhöhen versuchen

Kay Hinze, Sebastian Sünkler und Dirk Lewandowski

Inhalt

1 Einleitung	2
2 Forschungsstand und Theoretischer Hintergrund	3
3 Untersuchungsmodell	10
4 Methodisches Vorgehen	10
5 Ergebnisse & Diskussion	14
6 Fazit	23
Anhang	25
Literatur	25

Zusammenfassung

Die Studie untersucht in einer Vollerhebung aller 6211 Kandidierenden zur Bundestagswahl 2021, inwieweit diese auf ihren persönlichen Websites im Wahlkampf Suchmaschinenoptimierung (SEO) angewandt haben. Dies ließ sich auf 93 Prozent der 1372 untersuchten Websites feststellen. Auf individueller Ebene der Kandidierenden zeigt sich eine Bedeutung der Professionalisierung für die Nutzung von SEO: Wer in den Bundestag eingezogen ist, nutzt eher SEO. Auf organisationsbezogener Ebene zeigt sich, dass eine Zugehörigkeit der Partei zum Parlament die SEO-Nutzung der Kandidierenden begünstigt.

Schlüsselwörter

Suchmaschinen · SEO · Google · Wahlkampf · Bundestagswahl

K. Hinze (✉)
ARGUS DATA INSIGHTS Deutschland GmbH, Berlin, Deutschland
E-Mail: mail@kayhinze.de
S. Sünkler · D. Lewandowski
Department Information, Hochschule für Angewandte Wissenschaften Hamburg, Hamburg, Deutschland
E-Mail: sebastian.suenkler@haw-hamburg.de, dirk.lewandowski@haw-hamburg.de

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2022
K.-H. Korte et al. (Hrsg.), *Die Bundestagswahl 2021*,
https://doi.org/10.1007/978-3-658-35758-0_19-1

https://link.springer.com/referenceworkentry/10.1007/978-3-658-35758-0_19-1

5. GEMEINSAME STUDIE



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

5. GEMEINSAME STUDIE: BRAINSTORMING

Haben Sie eine Idee, was Sie mit Hilfe des RAT untersuchen könnten?

- Bitte schreiben Sie Studienideen auf die Karten (eine Idee pro Karte).
- Schreiben Sie bitte den Studentyp (z. B. Retrievalstudie) und das übergeordnete Thema auf (z. B. Vergleich der Qualität eines Bibliothekssystems vs. Google)



5. GEMEINSAME STUDIE: EINSCHRÄNKUNGEN

Beschränkungen für Studien im Rahmen des Tutorials

- **Vorhandene** Funktionen
- Maximal **zwei Suchanfragen** und **zwei Suchmaschinen**
- Maximal **fünf Suchergebnisse** pro Suchmaschine
- "**Fast scraping**": Webseiten werden nur teilweise gespeichert



5. GEMEINSAME STUDIE: DURCHFÜHRUNG

The screenshot displays the RAT interface with a dark navigation bar at the top. The navigation bar includes a logo, 'Dashboard', 'New Study', 'Home', 'About', 'Features', 'Contact', and a user profile 'sebastian.suenkler@haw-hamburg.de'. Below the navigation bar, a progress indicator shows four steps: 1. STUDY DETAILS (active), 2. SEARCH ENGINES, 3. SEARCH QUERIES, and 4. RESULT OPTIONS. The main content area is titled 'STUDY DETAILS' and contains three input fields: 'Study name' (a text box), 'Description' (a larger text area), and 'Study type' (a dropdown menu currently showing 'Relevance Assessment'). A 'Next >>' button is located at the bottom right of the form.



6. ABSCHLUSS



112.BIBLIO CON ²⁰/₂₄
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

**SEARCH
STUDIES**



**HAW
HAMBURG**

6. ABSCHLUSS

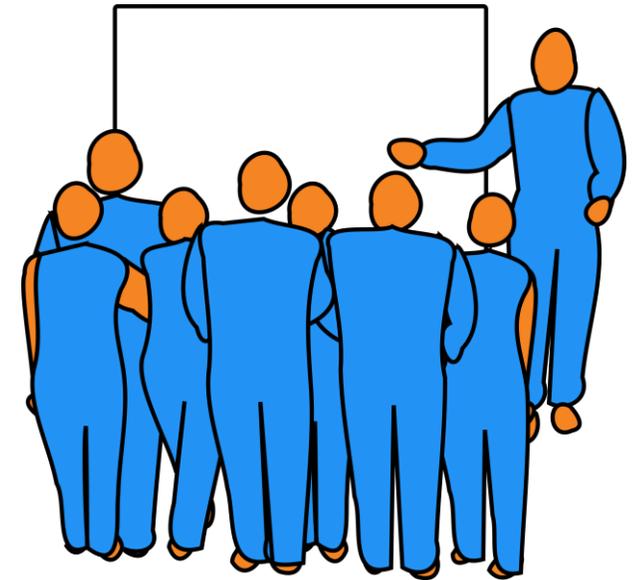
Was werden Sie in diesem Tutorial lernen?

Sie haben den **Umgang mit RAT** und seine technischen Möglichkeiten kennen gelernt.

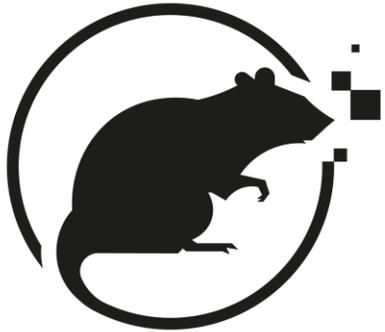
Was werden wir heute gemeinsam erleben?

Wir haben eine Studie an RAT **konzipiert** und **durchgeführt**.

Außerdem haben Sie jetzt **Ideen gewonnen**, wie Sie das RAT in Zukunft einsetzen können.



6. ABSCHLUSS: KONTAKT



RAT
RESULT ASSESSMENT TOOL

Projekt-Website: <https://searchstudies.org/research/rat/>

Demo: <https://rat-software.org/>

Kontakt für Studien: rat@searchstudies.org

Call for Studies: <https://searchstudies.org/call-for-studies/>

RAT
COMMUNITY
MEETING **2024**

Friday
27th Sept.

Hamburg,
Germany

Save the date!

<https://searchstudies.org/community-meeting>

26



112.BIBLIO CON 2024
offen.lokal.global.
04. - 07.06.2024 | Hamburg



RAT
RESULT ASSESSMENT TOOL

SEARCH
STUDIES

HAW
HAMBURG

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT



Sebastian Sünkler

sebastian.suenkler@haw-hamburg.de

<https://searchstudies.org/team/sebastian-suenkler/>

Danksagung:

Dieses Forschungsprojekt "RAT" wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert (Förderkennzeichen 460676551).



112.BIBLIO CON 20
24
offen.lokal.global.
04. - 07.06.2024 | Hamburg



**SEARCH
STUDIES**

**HAW
HAMBURG**