Hamburg, October 10, 2024

Marco Piangerelli,
PeerJ Computer Science

Dear Marco Piangerelli,

We want to thank the reviewers for their constructive comments, all of which we were happy to implement. In the attached table, you will find our answers to the reviewers' comments, including the line numbers where the changes can be found in the manuscript.

We hope that the manuscript is now suitable for publication in *PeerJ Computer Science*.

Sincerely,

Sebastian Sünkler M.A.
Hamburg University of Applied Sciences, Hamburg, Germany
E-Mail: sebastian.suenkler@haw-hamburg.de

On behalf of all authors.

Response to the referees' comments for PeerJ Computer Science paper:
"Result Assessment Tool (RAT): Empowering search engine data analysis"

Reviewer 1 (Valeria Mazzeo)			
Code	**Comment**	**Response** ***Please note: All line numbers refer to the line numbers in the review PDF of the revision***	**Reviewer's comment**
R1_1	In the abstract, could you please clarify the study participants? If there are character constraints, I would suggest removing the explicit reference to Jupyter Notebook, as it may not contribute significantly to the understanding of the study's use.	We removed the explicit reference to Jupyter Notebook.	Thank you.

R1_2	When examining the field of health, for example, we found studies conducting quality evaluations (e.g., Janssen et al., 2018) or content analyses (e.g., Rachul et al., 2020) of health- related search results, which are particularly relevant given the increasing importance of assessing online health information for accuracy and reliability in guiding public health decisions. Overall, this passage lacks detail on how RAT could be implemented in practice. For example, it would be helpful to explain how RAT could be adapted for studies in fields like health and media, including any specific modifications or applications that would make it suitable for these areas of research.	Thank you for bringing this to our attention. To prevent excessive detail regarding the implementation of the RAT in the introduction, we have removed specific study details from the introduction and referenced the results section instead. In the results section, we have included more information about the use of the RAT. Lines 61-66 Lines 639-656	Thank you for the revisions. I appreciate the additional examples of RAT's potential applications. However, I believe more specificity is needed regarding how RAT can be adapted to address challenges such as bias. Please refer to my comments below.

R1_3	The text mentions ideological bias in search results, but it does not directly tie this to the application of RAT. This point feels a bit disconnected. Could you please explain how RAT could help identify or mitigate these biases?	We added an explanation of how RAT can be used for such studies such as Ballatore (2015). Lines 644-655	Thank you for adding further details on how RAT supports the classification of search results. However, I believe it could be more explicitly connected to my original concern. Specifically, how does RAT improve upon previous methods for classification? Additionally, I noticed a few points that need attention: - the accessed date for Ballatore's webpage is missing; - in the references throughout the paper, there seems to be no corresponding reference number in the bibliography. Some parts appear to be repeated, such as the reference to the small number of data analysed due to manual classification (see lines 71 and 98).
R1_4	Could you please elaborate on why access to larger datasets is so difficult or problematic? Are there any legal or technical barriers?	The barriers are now described in the text to explain why access to larger data sets is difficult. Lines 76-87	Thank you for the revision. I believe the new information adds value to the manuscript. However, I would suggest providing a clearer structure to the paragraph (e.g., breaking it into two sections, one focusing on RAT's applications in health and media, and another addressing the barriers to data access). As it stands, the paragraph feels quite dense. Additionally, while the paragraph briefly touches on RAT's potnetial to handle search engine result biases, it does not explicitly explain how RAT addresses these biases. This comment is related to my previous feedback linked to the previous one R1_2. It might also be helpful to refer to the section where this aspect is discussed in more detail.

R1_5	Additionally, the concept of "self-interest" could be made more explicit and more directly tied to the potential consequences for research transparency. While the term "self-interest" is mentioned, the relationship between the financial motivations of search engine companies and the withholding of data could be explained more clearly. Could you also provide evidence showing how the financial interests of search engine providers might influence the ranking or visibility of search results? For example, this could include the prioritization of paid advertisements or content from partners.	We have now provided a detailed explanation of why independent studies on the self-interests of search engine companies are important, as it is highly unlikely that the search engines will (conduct and) publish such studies themselves. Lines 87-95	Thank you for addressing my query. Although the mention of the European Commission (2017) is included, there is no link to their report in the text. Could you please add the reference to the source? Thank you.

R1_6	Finally, there is no explicit discussion of how RAT could directly improve or be integrated into existing studies. For instance, you could provide a comparison with traditional methods to highlight how RAT offers potential improvements.	The “Results” section now discusses how RAT could improve the existing studies. Lines 592-604	Thank you for addressing my comment. I believe this paragraph now more effectively highlights the advantages of RAT over traditional manual and basic web-scraping approaches, particularly in terms of efficiency, scalability, and data preservation. However, there are a few queries/comments on areas where the argument could be strengthened further: - how does automation affect data quality? Automated scraping may introduce biases, such as favouring certain query structures. Could you provide more insights into how RAT accounts for mitigates these potential issues? - many search engines impose rate limits or employ anti-bot measures that can hinder large-scale data collection. How does RAT navigate these barriers? This has not been fully addressed in R1_4. - what measures are in place to ensure that automated data collection is as reliable (or more reliable) than manual methods? While I understand RAT can collect thousands of results, sheer quantity does not necessarily equate to better research quality. Are there specific research scenarios where traditional methods might still be preferable (e.g., in cases where qualitative analysis is needed)?

R1_7	“ Even though software cannot remove this barrier, RAT makes the process of assessing results efficient by removing duplicates from the results, providing a user interface for study participants (…)”. Could you please provide more information regarding the study participants? While the groups of researchers and participants are discussed later in the text (in the 'User Journey in RAT' section), it would be helpful to clarify their roles and how they interact with RAT earlier in the document.	The introduction provides more information regarding study participants now. Lines 106-110	Thank you for providing more information regarding the study participants.

R1_8	The following statement is not clear: “We also resolved an issue where the scraper returned fewer results than specified—e.g., scraping only 24 results when the limit was set to 30.” Could you please provide more details on the issue? Specifically, how can scraping fewer results than expected represent a problem? For example, if fewer results are scraped than the specified limit, how might this impact the quality or accuracy of the data?	We added an explanation of how it could impact the quality of data in the section Software quality assurance. Lines 544-557	Thank you for addressing my query and providing a clear explanation of how the bug impacted the data collection. You mention that the bug was caused by the scraper failing to handle Google’s dynamic loading mechanism. Could you please provide more details on how the bug was specifically identified? For instnce, was this issue observed across all queries or only specific types? After fixing the bug, what steps did you take to validate the collected data? Did you test the scraper against the same queries to confirm that it now retrieves the full set of results? One observation, if I may: while it is common for search engines to return fewer results than requested, the issue may not always stem from simply having fewer matching documents. There are other factors that could contribute to search engines returning fewer results, including query specificity (i.e., limited content matching, search engine algorithms and filters (e.g., region-based or language preferences), rate limits or anti-bot measures (these measures might also reduce the number of results being returned if the scraping mechanism hits a limit or is blocked), dynamic content (e.g., infinte scrolling),… In addition to the dynamic loading bug, were there any other potential biases in data collection that you considered? For instance, search engines constantly evolve their algorithms. Have you considered the potential for algorithmic changes or anti-bot measures (CAPTCHAs) that may still affect data collection accuracy? In line 221, you stated that automated queries are avoided, and you refrain from using methods to circumvent limitations such as solving captchas. Could you clarify how you handle situations where CAPTCHAs or rate-limiting measures appear during scraping? How do you ensure that it does not compromise the ethical or methodological integrity of your data collection?

R1_9	Please avoid the use of contracted forms (for example, “does not” instead of “doesn’t”, 'it is' instead of 'it's')	All contracted forms were changed to full forms.	Thank you, I appreciate that.
R1_10	Could you explain here what the focus of the software is? “We have limited ourselves to studies using data from commercial search engines such as Google or Microsoft Bing and library search systems, as this is the focus of the software.	We added an explanation of the focus of the software here and justified it. Lines 573-576	Thank you for providing further explanation. I have a couple of points that need clarification: - are there any trade-offs when applying RAT to other types of search systems that differ structurally from traditional SERPs? It would be helpful to discuss any limitations or considerations when adapting RAT to other types of search interfaces. - could you please clarify what is meant by “systematic collection” (“ RAT addresses this gap by providing tools to systematically collect, store, and analyze results from any web-based search interface, with particular attention to the complex structure of search engines result pages (SERPs) and library catalogs”). Does this mean RAT captures raw HTML, extracts structured data, or enables real-time monitoring of changes in SERPs? An example of how RAT works with search engines could also help clarify this point.
Reviewer 2 (Anonymous)			
Code	**Comment**	**Response** ***Please note: All line numbers refer to the line numbers in the review PDF of the revision***	

R2_1	The authors have made the requested improvements to a large extent, but I know that giving url information in parentheses in the text does not comply with the journal format. First, the Web platform name is written and details of citation should be provided in the references section. Please check the journal format again.	We received confirmation from the support team that our URL placement style is correct.	
R2_2	Some terms should be verified e.g. analysis --> anlayzing	We have reviewed the terms and adjusted them accordingly so that everything is standardized.	