How to Cite a Web Ranking and Make it FAIR

Alessandro Lotta[0009−0009−5047−606X] and
Gianmaria Silvello[0000−0003−4970−4554]

University of Padua, Padua, Italy
`<name.surname>@unipd.it`

Abstract. Citing data is crucial for acknowledging and recognizing the contributions of experts, scientists, and institutions in creating and maintaining high-quality datasets. It ensures proper attribution and supports reproducibility in scientific research. While data citation methods have focused on structured or semi-structured datasets, there is a need to address the citation of web rankings. Web rankings are significant in scientific literature, information articles, and decision-making processes. However, citing web rankings presents challenges due to their dynamic nature. In response, we introduce a new ”citation ranking” model and the *Unipd Ranking Citation tool*, designed to generate persistent and machine-readable citations, enhancing reproducibility and accountability in scientific research and general contexts. It is a user-friendly, open-source Chrome extension that employs ontology and RDF graphs for machine understanding and future reconstruction of rankings.

Keywords: Data citation · Ranking citation · Persistent citations

1 Introduction

Data citation has become a central topic in the scholarly domain and has a central role in science communication. Research on data citation has primarily revolved around two key aspects: establishing fundamental principles and developing architectural and computational solutions. Notably, two prominent international initiatives have been dedicated to defining the core principles for data citation. The first initiative, CODATA, published a comprehensive report on data citation principles in 2013 [1]. The second initiative, FORCE 11, presented a consolidated set of principles derived from various working groups in 2014 [12]. These principles underscore that data should be considered a research object worthy of citation, ensuring due recognition for data curators. Furthermore, they outline several key criteria that a citation should uphold, including:

- Enabling identification and access to the referenced data.
- Ensuring the persistence of data identifiers and associated metadata, addressing the issue of fixity.
- Guaranteeing the completeness of the reference, encompassing all necessary information for data interpretation and comprehension, even beyond the data’s lifespan.

- Promoting citation interoperability, allowing humans and machines to interpret and utilize the citations effectively.

Citing data is essential to acknowledge and recognize the contributions made by experts, scientists, and institutions who invest resources and expertise in creating, curating, and maintaining high-quality datasets. By citing data, we ensure that credit is properly attributed to those who deserve it. These datasets are crucial in conducting experiments, testing hypotheses, and advancing scientific knowledge. As their usage becomes more prevalent, it is crucial to acknowledge the efforts and dedication of those involved in producing and maintaining such valuable resources. Furthermore, data citation is vital in facilitating reproducibility in scientific research. We establish a permanent reference to the exact dataset or specific subset utilized in a series of experiments by including data citations. This ensures that others can easily locate and access the same data, enabling them to replicate the research findings and validate the results. Data citations serve as valuable pointers to the precise location of the data for reuse, making data more findable and promoting transparency and accountability in scientific investigations.

The primary emphasis in data citation methods has centered on citing structured or semi-structured datasets [24]. The aim has been to ensure the persistence of citations to specific portions of datasets [20], such as queries to relational [28] or graph databases [23]. Additionally, efforts have been directed toward enabling the retrieval of the exact same data being referenced over time [19]. Another critical aspect is ensuring the accuracy and comprehensiveness of the data citations, guaranteeing that they provide the necessary information to accurately locate and understand the referenced data. The applications of data citation encompass various domains, including the citation of CSV files, scientific centralized or federated databases, result tables generated by web applications, collections of objects obtained through interactive processes, and result sets derived from analytics methods.

Our research primarily focuses on addressing the overlooked aspect of citing web rankings. Web rankings are generated by web applications that utilize search engines to provide relevant data or documents in response to specific user queries. Typically, a user expresses their information needs through a keyword query, and the resulting ranking represents a list of potentially relevant objects for that query. Prominent examples of web rankings include those generated by web search engines like Google and Bing and academic search engines like Google Scholar or Scopus for literature searches. However, search engines are also employed by social networks like Twitter, which generate rankings of relevant tweets based on specific hashtags or keywords. Web rankings play a significant role in scientific literature. For instance, researchers may utilize web rankings to illustrate previous studies' absence by searching on platforms such as Google Scholar or PubMed. They may also present a collection of relevant tweets on a trending societal topic to provide context and motivation for a study. Additionally, web rankings can support decision-making processes by showcasing the results of a patent search on a specialized search engine.

We introduce “citation ranking,” a model and an open tool designed to generate FAIR (Findable, Accessible, Interoperable, and Reusable) citations for web rankings. The main challenge we address is the creation of persistent, human- and machine-readable citations for web rankings, which are inherently dynamic and subject to change due to various factors, including user preferences and contextual settings. With “citation ranking,” we aim to enable stable referencing to transient web rankings. Currently, it is not feasible to mention a specific ranking, such as papers, web pages, or tweets, and allow third parties, including researchers and the general public, to reproduce and verify the existence of that specific ranking. This poses a significant obstacle to reproducibility and accountability in scientific research and general information articles where web rankings are frequently cited as evidence.

We provide a user-friendly tool that ensures web rankings can be treated as stable and citable objects: the *Unipd Ranking Citation tool*. By doing so, we aim to promote reproducibility and accountability in scientific endeavors and general contexts where web rankings are utilized as evidence. The ultimate goal is to enhance the reliability and transparency of information derived from web rankings, fostering a more robust and trustworthy knowledge ecosystem.

This work provides the first free-to-use and open-source tool to create FAIR and persistent citations of Web rankings. The ranking citation tool is provided as a Chrome plug-in/extension easily usable from a commonly employed browser. We provide a citation model for Web rankings, including human- and machine-readable serializations of the ranking to be cited. To this end, we defined an ontology to create machine-readable Resource Description Framework (RDF) graphs serializing the ranking, enabling inference, machine-understanding, and the reconstruction of the ranking for future purposes. Currently, the *Unipd Ranking Citation tool* works for Google Scholar, Google, Bing, Scopus, and Twitter.

The rest of the paper is organized as follows: Section 2 overviews state of the art in data citation, reporting the necessity to cite Web rankings and the absence of viable solutions. Section 3 presents the citation model for Web rankings. Section 4 details the Unipd Ranking Citation tool technical architectures explaining how it has been implemented as an extension of Chrome. Section 5 describes a use case based on Google Scholar. Finally, Section 6 draws some final remarks.

2 Background

Within the Research Data Alliance (RDA) initiative, two working groups specifically address the topic of data citation. The first is the Data Citation Working Group (WG), [1] which focuses on establishing methodologies for persistently citing subsets of data derived from queries to structured databases. It aims to develop approaches that enable accurate and traceable referencing of specific data portions obtained through querying structured databases.

[1] `https://www.rd-alliance.org/groups/data-citation-wg.html` [visited on 22 May 2023]

The second working group is the Complex Citation WG, [2] which concentrates on the citation and distribution of credit for extensive collections of objects. Their focus extends beyond individual data subsets and encompasses the citation practices and mechanisms for acknowledging and attributing credit to large-scale collections of diverse objects. The objective is to devise methods that facilitate proper citation and recognition for researchers and contributors in creating and curating such extensive collections. Both working groups within the RDA initiative play crucial roles in advancing the field of data citation by addressing different aspects of citation methodology. By studying and providing solutions for persistent data subset citation and complex object collection citation, these groups contribute to establishing standardized practices that enhance traceability, reproducibility, and credit attribution in data-intensive research. The activities undertaken by these working groups do not specifically tackle the challenge of citing web rankings. However, it is worth noting that the Data Citation Working Group recognizes the citation of information retrieval rankings, such as those generated by search engines, as a critical issue to address for ensuring the reproducibility of scientific research [21]. To our knowledge, no viable solutions have been proposed to tackle the issue.

[24] provides an extensive overview of state of the art in data citation up to 2018, where the citation of web rankings is never mentioned. Over the past five years, there has been a notable increase in awareness regarding the significance of data citation, leading to the establishment of guidelines for citing datasets by many publishing houses (e.g., Springer Nature [15] and Elsevier).[3] Various domains, including neuroimaging [13], geoscience [2, 16], and biology [18, 22, 26], have explored the incorporation of data citation practices into their research outputs. Numerous studies have delved into the distribution of credit among large groups of scientists who contribute to datasets or data aggregations [7, 8, 11, 17, 27]. These works have proposed novel measures, introduced new authorship categories, and explored credit distribution mechanisms [9, 10]. Considerable efforts have also been invested in developing infrastructures for depositing datasets, ensuring comprehensive descriptions, and enhancing their discoverability and accessibility [5, 6].

Data citation in scholarly graphs has been recognized for its impact and importance. Efforts have been made to extend existing citation graphs to include data, enabling seamless integration of datasets [4]. Furthermore, studies have examined the relationship between datasets and scholarly papers in the scientific discourse, uncovering the connections between them [14]. These endeavors contribute to a more comprehensive understanding of research and facilitate the effective dissemination and utilization of data in scholarly communication [3].

However, despite these initiatives and advancements, none have explicitly targeted rankings' citations. While the importance of data citation has been

[2] `https://www.rd-alliance.org/groups/complex-citations-working-group` [visited on 22 May 2023]

[3] `https://www.elsevier.com/authors/tools-and-resources/research-data` [visited on 22 May 2023]

acknowledged and pursued in various disciplines, the specific challenge of citing web rankings remains unaddressed.

3 Citation Model

In data citation, two fundamental elements comprise a citation: the data object being referenced and the accompanying reference or citation snippet that describes the cited data. The data object must possess persistence, ensuring its continuous accessibility in the exact form as initially cited. Conversely, the reference should possess reusability, allowing machines and humans to interpret and utilize it effectively. Furthermore, the reference should conform to a consistent format observed by other citations referencing the same class of objects, ensuring correctness and completeness. Lastly, an essential characteristic of the reference is its ease of creation, avoiding the need for manual effort during the citation process.

The dynamic and transient nature of web rankings stems from their susceptibility to change based on factors such as the user initiating the query, the contextual circumstances surrounding it, and updates to the underlying index. Therefore, ensuring the longevity of web rankings requires storing them in a format that facilitates long-term preservation while simultaneously enabling machine interpretation and human comprehension.

To ensure human readability, we capture a screenshot(s) of the webpage(s) displaying the ranking to be cited in the PNG (Portable Network Graphics) format. The PNG format is a lossless compressed format widely recognized for its suitability in the long-term preservation of images. It is recommended by institutions such as the Library of Congress for its preservation qualities. [4]

To ensure machine readability, two main steps are taken. Firstly, essential information from the web ranking, including the title, description snippet, URL, position on the page, user, settings, and the main characteristics of the search engines, is extracted. This process involves capturing the key textual components that determine the ranking. This extracted information creates an RDF graph. The RDF graph is a structured representation of the extracted data, enabling machines to interpret and process the information effectively. The key textual elements forming the ranking can be reconstructed from the RDF graph, facilitating machine-based analysis and utilization of the ranking data. Of course, an external service or web application can employ the RDF graph to produce a human-readable replica of the original ranking.

To enhance the machine interpretability of the created RDF graph, we have developed a concise ontology, i.e., the Ranking Citation Ontology (RCO). This ontology serves the purpose of representing the specific domain of interest. Figure 1 reports the graphical representation of the RCO, publicly available at `https://rankingcitation.dei.unipd.it/ontology/`. We can see that the class `User`

[4] See `https://www.loc.gov/preservation/resources/rfs/stillimg.html` and `https://howtofair.dk/how-to-fair/file-formats/` [last visited on 24 May 2023].

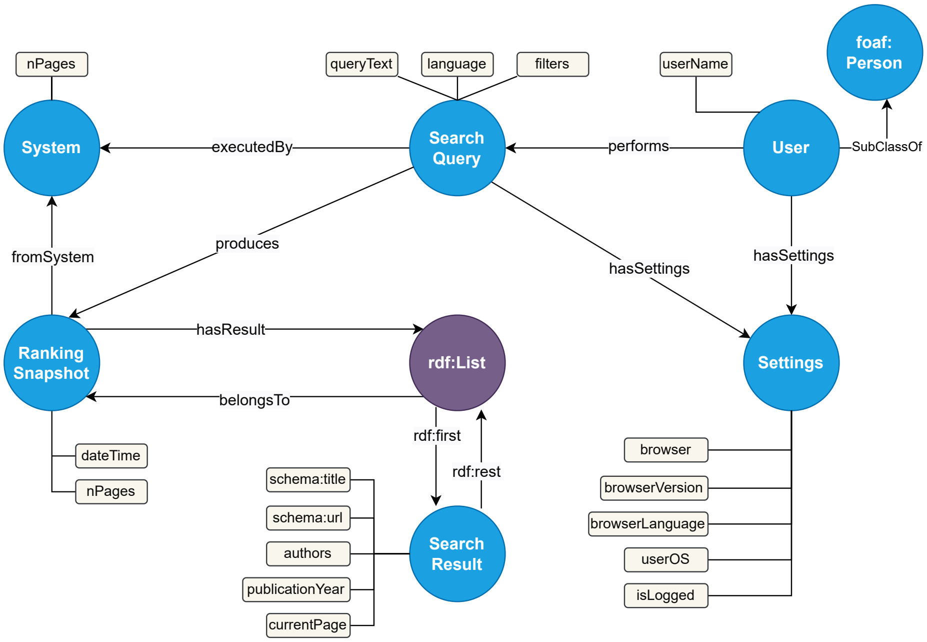


Fig. 1: A graphical representation of the Ranking Citation Ontology

models the user issuing a `Search Query` to a `System`. The key properties are the language of the query, the active filters (if any), the text of the query, and the number of result pages the system displayed. The `Search Query` produces a `Ranking Snapshot`, which is what we capture (the number of pages captured is a user setting). The `Ranking Snapshot` is composed of a list (i.e., an `RDF List`) of `Search Results`. A `Search Result` comprises several properties such as the title, the URL, and the current page meaning in which web page the result is displayed. Moreover, we also store the authors and the publication year for search systems like Google Scholar, where a search result corresponds to a scientific paper. Finally, we represent the user and system `Settings` such as the browser type, version, language, operating system, and if the user was logged in when performing the search.

In the final step, we package the citation artifact using the Research Object (RO) Crate [25]. RO Crate is an openly developed specification offering a lightweight and adaptable packaging format for research objects. It is a structured container encompassing research data, metadata, and contextual information to ensure their integrity, provenance, and discoverability. The format relies on JSON-LD with `schema.org` annotations, providing a means for data persistence and ensuring long-term accessibility. The RO Crate ontology defines the vocabulary and relationships utilized to describe the contents within an RO Crate. We employ RO Crate to describe the objects stored to preserve a web ranking, associate the screenshot images with the RDF graph, and makes the

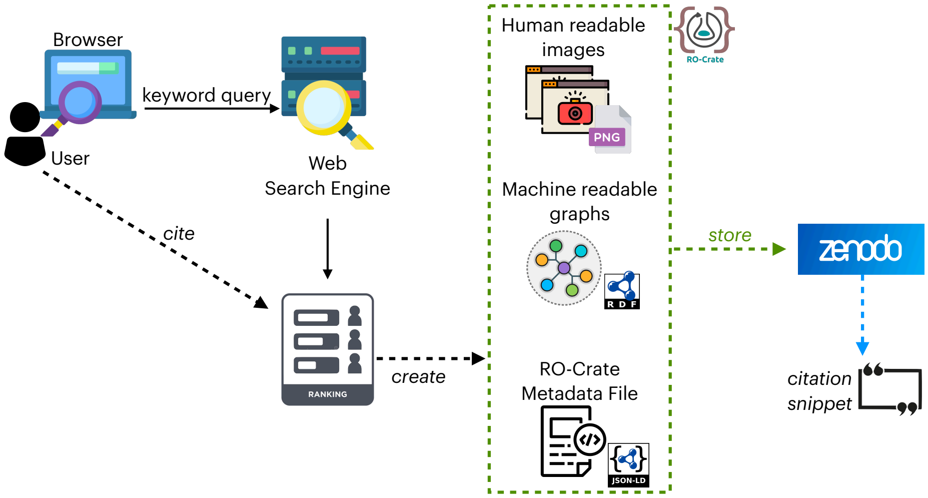


Fig. 2: The ranking citation pipeline.

entire citation bundle interpretable. Additionally, the RO Crate contains metadata related to the user, such as their name, ORCID, and affiliation, enabling appropriate attribution of generated citations to the respective user and institution. To facilitate data deposition and guarantee long-term accessibility with robust preservation practices, we combine RO Crate with Zenodo. This integration allows for the seamless deposition of the citation bundle while ensuring enduring accessibility and preservation of the data.

Figure 2 illustrates the key components of the ranking citation model. The process begins with a user issuing a query to a search engine, which generates a ranking. When the user requests a citation for the ranking, three distinct objects are generated and bundled within an RO Crate. These objects include the screenshot images, the RDF graph, and the RO-Crate metadata file. The RO Crate, containing these objects, is then securely stored in Zenodo for long-term preservation, ensuring the persistence and accessibility of the citation. As a result, a consistent citation snippet can be generated, allowing for proper referencing of the ranking.

4 Architecture and Implementation of the Ranking Citation Tool

We developed the proposed model as a Chrome plugin/extension, seamlessly integrating it into a browser for easy use by stakeholders. The ”Unipd Ranking Citation Tool” plugin was built using the Chrome Extension CLI development structure. This framework provides a predefined structure with essential folders and source files. The ”src” folder contains the background script, content script, popup script, and stylesheets for HTML pages. The ”public” folder includes

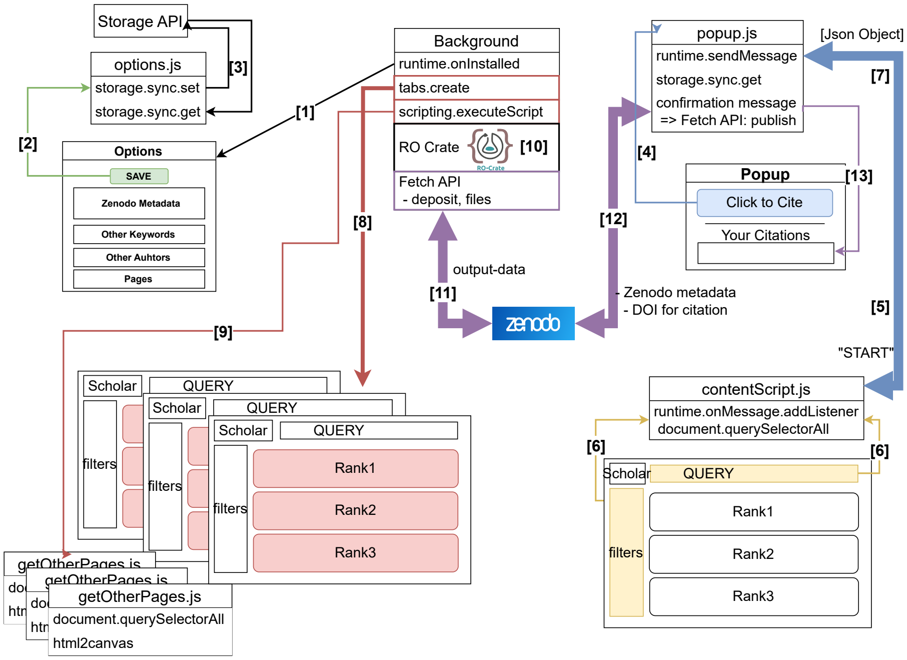
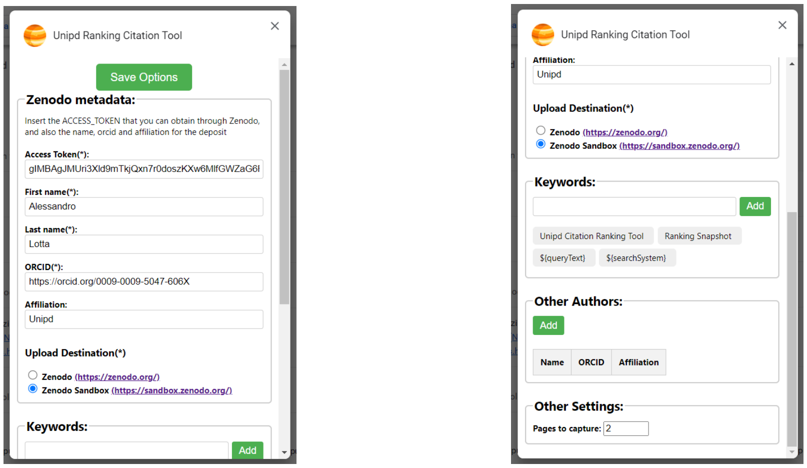


Fig. 3: The *Unipd Ranking Citation Tool* architecture diagram

HTML files, including the code for the options page. It also houses subdirectories for storing the required icons and the vital "manifest.json" file. This file contains crucial information about the extension, such as its name, version, permissions, and declared scripts, enabling proper loading and execution in the browser. The Chrome Extension CLI also configures Webpack by providing the necessary configuration files. This integration enables quick and simple development with an automatic reload feature, ensuring that any code changes are immediately reflected in the extension. Furthermore, it simplifies the compilation and packaging process of the extension. The 'build' folder is continuously updated throughout the development process to contain all the finalized files required for using and testing the extension in Chrome. This folder encapsulates the compiled and packaged extension, ready for deployment. The Chrome Extension CLI provides access to Node.js and the Node Package Manager (NPM) for efficient dependency management. This integration enables the easy inclusion and management of external libraries or frameworks.

In Figure 3 we can see the main components of the *Unipd Ranking Citation Tool* and how they interact. After installing the tool, the background script activates the `onInstalled` listener (step [1] in Figure 3). This listener triggers the `openOptionsPage` function, which directs the user to the options page specified in the manifest file (step [2]). On this page, the user can configure the settings



Unipd Ranking Citation Tool

Zenodo metadata:

Insert the ACCESS_TOKEN that you can obtain through Zenodo, and also the name, orcid and affiliation for the deposit

Access Token(*): gIMBAgJMUri3Xld9mTkjQxn7r0doszKXw6MlfGWZaG6l

First name(*): Alessandro

Last name(*): Lotta

ORCID(*): https://orcid.org/0009-0009-5047-606X

Affiliation: Unipd

Upload Destination(*)

○ Zenodo (https://zenodo.org/)
● Zenodo Sandbox (https://sandbox.zenodo.org/)

Keywords:

(a)

Unipd Ranking Citation Tool

Affiliation: Unipd

Upload Destination(*)

○ Zenodo (https://zenodo.org/)
● Zenodo Sandbox (https://sandbox.zenodo.org/)

Keywords:

Unipd Citation Ranking Tool | Ranking Snapshot | ${queryText} | ${searchSystem}

Other Authors:

Name	ORCID	Affiliation

Other Settings:

Pages to capture: 2

(b)

Fig. 4: Unipd Ranking Citation Tool: Setting page.

of the extension. In Figure 4, the main settings are displayed, including the Zenodo or Zenodo Sandbox account details, username, and ORCID. The user can choose the Zenodo sandbox for creating temporary or trial citations, or the Zenodo real instance for permanent citations. There are additional sections where users can add keywords for deposit metadata and specify additional authors or collaborators for the project/research. The user must input the respective individuals' names and ORCID for these sections. Finally, a section is dedicated to selecting the desired number of pages to capture during the research process.

After filling in all the required input fields, the user can save the settings by clicking the corresponding button. This action triggers a callback function that utilizes the integrated Chrome Storage API (step [3]) to save the data. The Storage API provides functions for asynchronous data manipulation, such as setting, updating, retrieving, and deleting data, specifically tailored for extensions. Our tool utilizes the "chrome storage sync" area, which synchronizes the data across all Chrome browsers where the user is logged in. If syncing is disabled, it behaves similarly to "storage.local," meaning that the data is cleared when the extension is removed. Using the "chrome storage sync set" function, the saved data is automatically populated in the input fields whenever the options page is reopened, enabling user editing.

Once the settings configuration is complete, users can access the extension's popup, which first checks if the current page URL is supported by the tool (see step [4] in Figure 3). If not supported, a message is displayed indicating that citations are unavailable on the current page. Below this message, the "Your Citations" section appears, displaying a list of citation cards from previous captures. The "Your Citations" section remains visible regardless of the visited site.

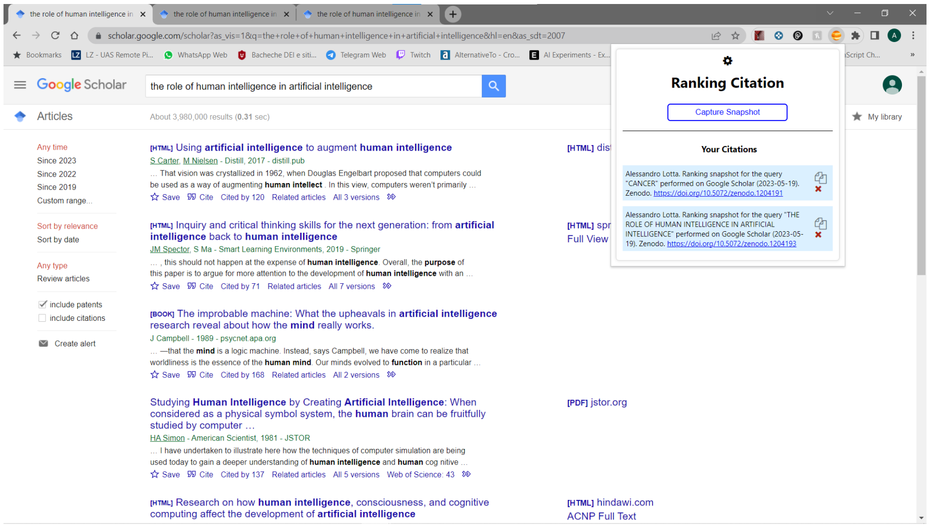
If the user opens the extension on a compatible page, the popup displays a button for capturing data. When clicked, the popup script sends a message to the content script injected into the currently viewed page (see step [13]). The browser's message-passing framework facilitates communication between these scripts. In this case, a one-time JSON-serializable message is sent using the "runtime.sendMessage" function, which includes information about the active page. On the receiving end, the content script implements a "runtime.onMessage" listener to capture any message containing the keyword 'START' (see [5]). Upon receiving such a message, the content script captures the required data from the result page (see [6]).

The content script initially defines the RDF graph's necessary classes, data, and object properties. It then analyzes the page's Document Object Model (DOM) to extract data related to the SearchQuery, System, RankingSnapshot, Settings, and User classes. After collecting the required data, the content script adds the individuals to a JavaScript object that will compose the graph. Finally, the content script sends a response message containing the RDF graph stored as a JSON object back to the popup (see step [7]).

After receiving the content's response, the popup initiates a new simple one-time request to communicate with the background script, sending the received data as the payload. The background script receives the message and opens multiple new pages based on the extension's options settings. The tool utilizes the "chrome.tabs.create" function from the integrated "chrome.tabs" API to create these new pages. It is important to note two aspects in this process: firstly, the filters set during the search process are maintained on the newly opened pages, ensuring consistency. Secondly, a new script is injected into each opened page using the "chrome.scripting.executeScript" function. These injected JavaScript files are responsible for gathering the remaining data necessary for ranking the results. They employ a similar approach to scrape the DOM as described earlier.

The captured ranks consider both the "currentPage" parameter, indicating the page where they are found, and the order assigned by the ranking. Each injected script sends the collected data back to the background script through a one-time request. The background script waits until all the scripts have been completed before proceeding. At this stage, the tool enters the upload phase (step [12]). In the first phase, an RO Crate is created by defining a JSON object that encapsulates all the entities within the Crate. This object includes the context and the graph representing the generated output files, ensuring proper organization within the deposit. Subsequently, the JSON object containing the gathered data and the RO Crate are converted into JavaScript File variables, preparing them for publication.

Next, the deposited metadata is defined, including the title, notes, description, keywords, and authors specified in the options. The files are sent to the server using the JavaScript Fetch API and its asynchronous function "fetch". This step involves three consecutive fetch calls: one for creating the deposit in Zenodo and two for uploading the two files. If the deposit creation is successful,



the role of human intelligence in artificial intelligence

About 3,980,000 results (0.31 sec)

[HTML] Using artificial intelligence to augment human intelligence
S Carter, M Nielsen - Distill, 2017 - distill.pub
… That vision was crystallized in 1962, when Douglas Engelbart proposed that computers could be used as a way of augmenting human intellect . In this view, computers weren't primarily …
Save Cite Cited by 120 Related articles All 3 versions

[HTML] Inquiry and critical thinking skills for the next generation: from artificial intelligence back to human intelligence
JM Spector, S Ma - Smart Learning Environments, 2019 - Springer
… , this should not happen at the expense of human intelligence. Overall, the purpose of this paper is to argue for more attention to the development of human intelligence with an …
Save Cite Cited by 71 Related articles All 7 versions

[BOOK] The improbable machine: What the upheavals in artificial intelligence research reveal about how the mind really works.
J Campbell - 1989 - psycnet.apa.org
… —that the mind is a logic machine. Instead, says Campbell, we have come to realize that worldliness is the essence of the human mind. Our minds evolved to function in a particular …
Save Cite Cited by 168 Related articles All 2 versions

Studying Human Intelligence by Creating Artificial Intelligence: When considered as a physical symbol system, the human brain can be fruitfully studied by computer …
HA Simon - American Scientist, 1981 - JSTOR
… I have undertaken to illustrate here how the techniques of computer simulation are being used today to gain a deeper understanding of human intelligence and human cog nitive …
Save Cite Cited by 137 Related articles All 5 versions Web of Science: 43

[HTML] Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence

[HTML] dist
[HTML] spr
Full View
[PDF] jstor.org
[HTML] hindawi.com
ACNP Full Text

Ranking Citation

Capture Snapshot

Your Citations

Alessandro Lotta. Ranking snapshot for the query "CANCER" performed on Google Scholar (2023-05-19). Zenodo. https://doi.org/10.5072/zenodo.1204191

Alessandro Lotta. Ranking snapshot for the query "THE ROLE OF HUMAN INTELLIGENCE IN ARTIFICIAL INTELLIGENCE" performed on Google Scholar (2023-05-19). Zenodo. https://doi.org/10.5072/zenodo.1204193

Fig. 5: "Your Citations" as displayed by the Chrome extension tool. The user can copy and paste the automatically created citation snippet pointing to a FAIR and persistent citation.

the service responds with the deposit ID, which is necessary for the subsequent uploading process.

Finally, the background script sends a message to the scripts of the opened pages that were used to capture the rankings instructing them to capture a screenshot of each page (step [8]). This passage is executed using the *html2canvas* library[5]. *html2canvas* takes the HTML document's body as input and returns a canvas element representing the entire visible page. The canvas is converted into a blob and subsequently into a file variable, uploaded to the same deposit using a "fetch" call. The scripts injected on each page notify the popup that the screenshots are taken by sending a message.

The final step performed by the extension occurs in the popup script, which prompts the user to confirm the publishing of the deposit to Zenodo or Zenodo Sandbox. Upon confirmation, the popup initiates a publish request to the designated upload destination using a fetch call. The response returned by the service is used to construct the citation text on the "Your Citations" section of the popup.

5 Use Case: Google Scholar

In this use case scenario, we will walk through the process of using the *Unipd Ranking Citation Tool* Chrome extension. To begin, users can update their existing extension or install it from the Chrome Web Store under the name *Unipd*

[5] `https://html2canvas.hertzen.com/` [last visited on 30 May 2023.]

Ranking Citation Tool. Open a new tab in Google Scholar and enter a query to search for relevant literature. Once the search results are displayed, you can access the extension by clicking on its icon in the browser's top right corner. The extension will present a button indicating the availability of citations on the current page, along with a list of previously published citations as shown in Figure 5.

By clicking the button, the extension will execute the necessary code in the background script and open a predefined number of new pages (as defined in the options). These pages will gather the data for the rankings on the search results page. The extension will display a confirmation message indicating that the file upload is complete and prompt the user to proceed with publishing on either Zenodo or its Sandbox. After confirming the publishing action, the extension's popup will proceed with the publication process. Once completed, a new card will be displayed, containing the citation for the deposit, Figure 5. Users can now navigate to Zenodo or its Sandbox and access the upload section to view the deposit. Clicking on the deposit will provide more details about the files contained within, including access to different versions, if available.

6 Final Remarks

In summary, our research introduces a novel solution for the citation of web rankings with the development of the *Unipd Ranking Citation tool*. This tool, available as a free and open-source Chrome plugin, addresses the need for FAIR and persistent citations of web rankings. This tool allows users to generate consistent and reliable citation snippets for web rankings, ensuring proper attribution and facilitating reproducibility in scientific research and other contexts.

The *Unipd Ranking Citation tool* represents a significant advancement in the field as it is the first model and tool specifically designed to address the challenges associated with citing web rankings. Currently, the tool is compatible with popular platforms such as Google Scholar, Google, and Twitter. However, our plans involve expanding its functionality to include other widely used rankings in research.

It is important to note that the tool relies on parsing the DOM of web pages, and as a result, it is subject to limitations. If the web page's structure being cited changes, the tool's parser may require updates to maintain its functionality. Nonetheless, the *Unipd Ranking Citation tool* provides a viable and practical solution for improving the citation of web rankings, promoting transparency and accountability in scientific research and beyond.

Resources

GitHub Repository: `https://github.com/aleLotta/ranking-citation.git`
Unipd Ranking Citation Tool: `https://rankingcitation.dei.unipd.it/`

Bibliography

[1] *Out of Cite, Out of Mind: The Current State of Practice, Polocy, and Technology for the Citation of Data*, volume 12. CODATA-ICSTI Task Group on Data Citation Standards and Practices, September 2013.

[2] Samuel C. Boone, Hayden Dalton, Alexander Prent, Fabian Kohlmann, Moritz Theile, Yoann Gréau, Guillaume Florin, Wayne Noble, Sally-Ann Hodgekiss, Bryant Ware, David Phillips, Barry Kohn, Suzanne O'Reilly, Andrew Gleadow, Brent McInnes, and Tim Rawling. Ausgeochem: An open platform for geochemical data preservation, dissemination and synthesis. *Geostandards and Geoanalytical Research*, 46(2):245–259, 2022.

[3] P. Buneman, G. Christie, J. A. Davies, S. D. Dimitrellou, R. andHarding, A. J. Pawson, J. L. Sharman, and Y. Wu. Why data citation isn't working, and what to do about it. *Database J. Biol. Databases Curation*, 2020, 2020.

[4] P. Buneman, D. Dosso, M. Lissandrini, and G. Silvello. Data citation and the citation graph. *Quant. Sci. Stud.*, 2(4):1399–1422, 2021.

[5] Adrian Burton, Amir Aryani, Hylke Koers, Paolo Manghi, Sandro La Bruzzo, Markus Stocker, Michael Diepenbroek, Uwe Schindler, and Martin Fenner. The scholix framework for interoperability in data-literature information exchange. *D Lib Mag.*, 23(1/2), 2017.

[6] H. Cousijn, P. Feeney, D. Lowenberg, E. Presani, and N. Simons. Bringing citations and usage metrics together to make data count. *Data Science Journal*, 18(1):9, 2019.

[7] Mark R. Cullen, Michael Baiocchi, Lisa Chamberlain, Isabella Chu, Ralph I. Horwitz, Michelle Mello, Amy O'Hara, and Sam Roosz. Population health science as a unifying foundation for translational clinical and public health research. *SSM - Population Health*, 18:101047, 2022.

[8] Thijs Devriendt, Mahsa Shabani, and Pascal Borry. Data sharing in biomedical sciences: A systematic review of incentives. *Biopreservation and Biobanking*, 19(3):219–227, 2021. PMID: 33926229.

[9] D. Dosso, S. B. Davidson, and G. Silvello. Credit distribution in relational scientific databases. *Inf. Syst.*, 109:102060, 2022.

[10] D. Dosso and G. Silvello. Data credit distribution: A new method to estimate databases impact. *Journal of Informetrics*, 14(4):101080, 2020.

[11] Robert M. Ewers, Jos Barlow, Cristina Banks-Leite, and Carsten Rahbek. Separate authorship categories to recognize data collectors and code developers. *Nature Ecology & Evolution*, 3(12):1610–1610, 2019.

[12] FORCE-11. *Data Citation Synthesis Group: Joint Declaration of Data Citation Principles.* FORCE11, San Diego, CA, USA, 2014.

[13] Corey Horien, Stephanie Noble, Abigail S. Greene, Kangjoo Lee, Daniel S. Barron, Siyuan Gao, David O'Connor, Mehraveh Salehi, Javid Dadashkarimi, Xilin Shen, Evelyn M. R. Lake, R. Todd Constable, and Dustin Scheinost. A hitchhiker's guide to working with large, open-source neuroimaging datasets. *Nature Human Behaviour*, 5(2):185–193, 2021.

[14] O. Irrera, A. Mannocci, P. Manghi, and G. Silvello. A Novel Curated Scholarly Graph Connecting Textual and Data Publications. *ournal of Data and Information Quality*, in print, 2023.

[15] V. Khodiyar. The basics of data citation. URL: `https://researchdata.springernature.com/posts/the-basics-of-data-citation`, May 2021.

[16] Xin Li, Guodong Cheng, Liangxu Wang, Juanle Wang, Youhua Ran, Tao Che, Guoqing Li, Honglin He, Qiang Zhang, Xiaoyi Jiang, Ziming Zou, and Guofeng Zhao. Boosting geoscience data sharing in china. *Nature Geoscience*, 14(8):541–542, 2021.

[17] P. Mongeon, N. Robinson-Garcia, W. Jeng, and R. Costas. "incorporating data sharing to the reward system of science: Linking datacite records to authors in the web of science". *Aslib Journal of Information Management*, 69(5):545–556, 2017.

[18] V. H. Oza, J. H. Whitlock, E. J. Wilk, A. Uno Antonison, B. Wilk, M. Gajapathy, T. C. Howton, A. Trull, L. Ianov, E. A. Worthey, and B. n. Lasseigne. Ten simple rules for using public biological data for your research. *PLoS Comput Biol*, 19(1):e1010749, 2023.

[19] S. Pröll and A. Rauber. A Scalable Framework for Dynamic Data Citation of Arbitrary Structured Data. In *Proc. of 3rd Int. Conf. on Data Management Technologies and Applications*, pages 223–230, 2014.

[20] A. Rauber, A. Ari, D. van Uytvanck, and S. Pröll. Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. *Bulletin of IEEE Technical Committee on Digital Libraries, Special Issue on Data Citation*, 12(1):6–15, May 2016.

[21] A. Rauber and M. Parsons. Data Citation Working Group Mtg @ P19. URL: `https://www.rd-alliance.org/system/files/documents/220623_rda_p19_wgdc_slides.pdf`, slide 52, June 2022.

[22] Katharina Sielemann, Alenka Hafner, and Boas Pucker. The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ*, 8:e9954, September 2020.

[23] G. Silvello. A Methodology for Citing Linked Open Data Subsets. *D-Lib Magazine*, 21(1/2), 2015.

[24] G. Silvello. Theory and Practice of Data Citation. *Journal of the American Society for Information Science and Technology (JASIST)*, 69(1):6–20, 2018.

[25] S. Soiland-Reyes, P. Sefton, M. Crosas, L. J. Castro, F. Coppens, J.. Fernández, D. Garijo, B. Grüning, M. La Rosa, S. Leo, E. ÓCarragáin, M. Portier, A. Trisovic, RO-Crate Community, P. Groth, and C. Goble. Packaging research artefacts with ro-crate. *Data Science*, 5(2):97–138, 2022.

[26] Paul Villoutreix. What machine learning can do for developmental biology. *Development*, 148(1), 01 2021. dev188474.

[27] Mark Westoby, Daniel S. Falster, and Julian Schrader. Motivating data contributions via a distinct career currency. *Proceedings of the Royal Society B: Biological Sciences*, 288(1946):20202830, 2021.

[28] Y. Wu, A. Alawini, S. B. Davidson, and G. Silvello. Data Citation: Giving Credit Where Credit is Due. In G. Das, C. M. Jermaine, and P. A. Bernstein,

editors, *Proc. of the 2018 International Conference on Management of Data, SIGMOD Conference 2018*, pages 99–114. ACM Press, New York, USA, 2018.