



There's an App for That

Tree of Science - ToS: A Web-based Tool for Scientific Literature Recommendation. Search Less, Research More!

Martha Zuluaga

Professor

Universidad Nacional Abierta y a Distancia

Core of Science

martha.zuluaga@unad.edu.co

Sebastian Robledo

Professor

Universidad Católica Luis Amigó

Core of Science

sebastian.robledogi@amigo.edu.co

Oscar Arbelaez-Echeverri

Researcher

Core of Science

technology@coreofscience.org

German A. Osorio-Zuluaga

Professor

Universidad Nacional de Colombia

gaosorioz@unal.edu.co

Nestor Duque-Méndez

Professor

Universidad Nacional de Colombia

ndduqueme@unal.edu.co

Abstract

Tree of Science (ToS) is an integrated web-based platform for a comprehensive analysis of scientific literature. ToS is designed to facilitate finding relevant literature and can be used by students, researchers, and academics. Based on graph theory metrics, this tool visualizes the works in a knowledge field as a tree where the roots are classic articles, the trunk represents those articles that allow the area to grow, and the leaves are the recently published articles. This article provides a description of how to search, format, and upload the data and identify significant literature in a specific research area. Finally, a brief description of how the Tree of Science works.

Keywords: Web-based tool, Recommender literature, Tree of Science, Scientific literature search, Scientometrics.

Recommended citation:

Zuluaga, M., Robledo, S., Arbelaez-Echeverri, O., Osorio-Zuluaga, G.A., & Duque-Méndez, N. (2022). Tree of Science - ToS: A web-based tool for scientific literature recommendation. Search less, research more! *Issues in Science and Technology Librarianship*, 100. <https://doi.org/10.29173/istl2696>

Introduction

Scientific literature reviews can be performed in a narrative mode or systematic procedure. However, in both cases, the researcher must minimize the bias to be critical, impartial, and avoid subjectivity. After identifying the research problem, the process is primarily concerned with selecting relevant literature available. This is a complex and time-consuming process due to the increased output of scientific publications. For example, the number of articles indexed in Web of Science (WoS) about systematic literature searching in 2015 is a hundred and fifty times greater than an identical search in 2001. In this regard, reviewing every paper published in a specific field becomes a difficult task. Therefore, it is necessary and valuable to develop informatic tools to facilitate the selection of relevant articles.

A systematic literature search requires building an equation (the topic question) with keywords and organizing phrases together with the Boolean operators ([Higgins et al., 2019](#); [Petticrew & Roberts, 2008](#)). This step usually requires five components to build a precise research question: the population, intervention, the comparator, the outcome or endpoint, and the study design ([Liberati et al., 2009](#)). The researcher then tries different arrangements until they find a precise combination that delivers relevant documents in the study area. They subsequently go through each article manually and evaluate the articles based on pre-established inclusion or exclusion criteria ([Liberati et al., 2009](#); [Pautasso, 2013](#)). This part must be done carefully to minimize risk of author bias.

Different tools have been developed to help visualize relationships between items in systematic search records. For example, Sci2 Tool ([Sci2 Team, 2009](#)), Bibexcel ([Persson et al., 2009](#)), CiteSpace II ([Chen, 2006](#)), VantagePoint ([Porter & Cunningham, 2004](#)), Network Workbench Tool ([NWB Team, 2006](#)), and databases such as Springer Link have been applying graph theory to show related articles. All listed tools allow the user to upload a data set in order create the network. Some tools include network visualization, others connect with open software for visualization like Gephi ([Bastian et al., 2009](#)) or Cytoscape ([Shannon et al., 2003](#)). However, these applications require knowledge of network processing and analysis.

One of the goals of our proposal, called *Tree of Science* (ToS), is to facilitate this process using graph theory, present the results as a tree, and recommend the articles by the position in the graph of citations. ToS developed a simple way to identify the knowledge area using a forest and tree metaphor. It is to lead the researcher to think about the scientific community where she or he would like to contribute (forest) and then think about the field application (tree).

The tree metaphor will help understand the field of research quickly, where the roots represent the classic articles, the trunk compiles the articles that made the field grow, and the leaves represent the newest articles that are highly connected with the trunk and roots. For this, we have developed ToS, a web-based tool for scientific literature analysis that facilitates the search task to non-programming users. ToS was specifically designed to address research and educational needs, including identification of relevant articles, linking the citation with the original page of the article, and presenting the results in an organic view to facilitate the classification ([Eggers et al., 2022](#)).

The limitation of ToS lies in the fact that the researcher may have access to the WoS Platform to download the query, as well as the access to databases to acquire the full paper.

ToS - Overview

The web-based tool can be accessed through this link: <https://tos.coreofscience.org/>. The home page presents the option to upload a WoS file in ToS (Figure 1) and the "Continue" button to process the file.



Figure 1. Tree of Science home page

Once the user sends the query downloaded from WoS, the system receives a .txt file retrieved by a pre-processing module that constructs the citation network and calculates the metrics to construct the results shown to the user (Figure 2). Since the application was developed for Web environments, the user interacts with the system through his or her preferred browser.

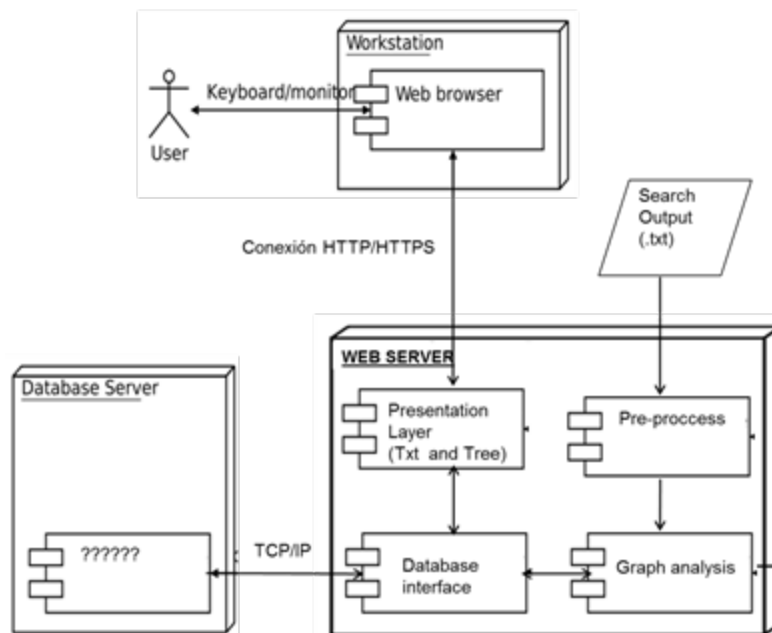


Figure 2. Deployment diagram

How ToS Works

ToS uses three algorithms to identify relevant articles, one for each indicator (root, trunk, and leaves). Roots represent the classical articles, or studies that started the knowledge area. The articles that connect classical articles with current articles (leaves) represent the trunk. Finally, leaves represent recently published papers and have many connections to the network. A detailed explanation of the algorithm could be found in the paper "SAP Algorithm for Citation Analysis: An improvement to Tree of Science" ([Valencia-Hernandez et al., 2020](#)). ToS has been applied to the literature analysis of Social Economy ([Duque et al., 2021](#)), medicine ([Gonzalez-Correa et al., 2021](#)), and Entrepreneurship ([Robledo et al., 2021](#)).

A Step-by-Step Guide Using a Study Case

In this section, a step-by-step guide protocol for using ToS is presented. In addition, how to search in WoS and how to download the references and citations is described. How to upload the data into ToS and interpret the tree is also included.

Develop a Search Equation: Forest and Tree Metaphor

ToS developed a simple way to identify the knowledge area using a forest and tree metaphor. The purpose is to lead researchers to determine where in a scientific subject area a contribution is situated (forest) and further to choose a field of application (tree). The equation should be created with the keywords that precisely integrate the information needed, and using the respective Boolean operators that articulate them. The ToS algorithm requires a minimum of 100 records from the search results to build the citation network. Using the tree analogy, if we have fewer than 100 records as a result of the search, we would see only one branch of the tree. However, the web tool has a file size limitation, which cannot exceed 5 MB. Based on our experience, this size is reached with around 500 records in the results. If you have more than 500 records in your results, you can process your tree using the open-source code "r-tos"

(<https://github.com/coreofscience/r-tos>). Additionally, you can also use the R package “tosr” (<https://cran.r-project.org/web/packages/tosr/index.html>).

Web of Science Search and Data Acquisition

In this example, the literature of scientometrics is analyzed. First, the search is conducted at WoS. Then, the user must select the Core Collection (Figure 3a); this is a critical step. Then, the user adds the keywords using the topic selected. The keyword "scientometrics" is used; it produced 7,440 entries. Using the metaphor explained before, the area of scientometrics is a forest. Then we have to look for just one "tree." For this, we added another keyword to specify which area of scientometrics we would like to search. We added the keyword: “algorithm*”, and it retrieved 285 results, see Figure 3b and Figure 3c.

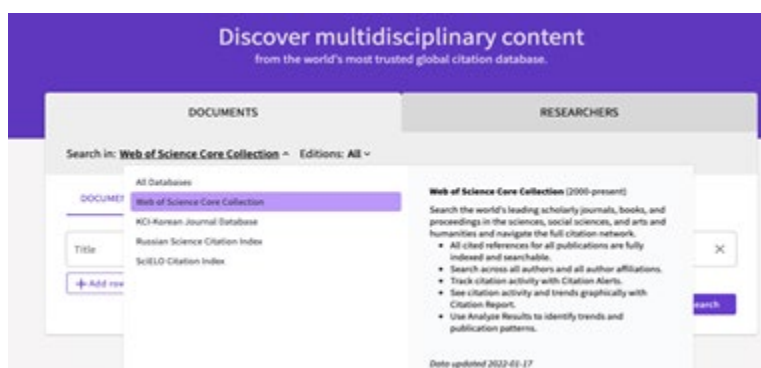


Figure 3a. Web of Science Core Collection

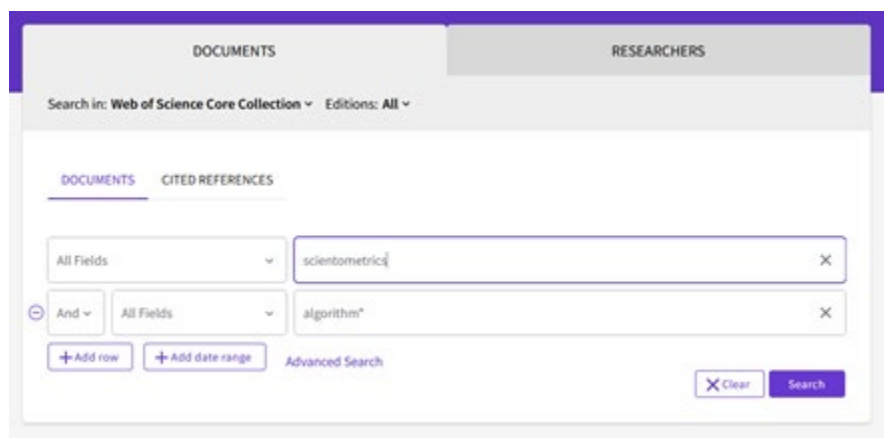


Figure 3b. Keywords and search question

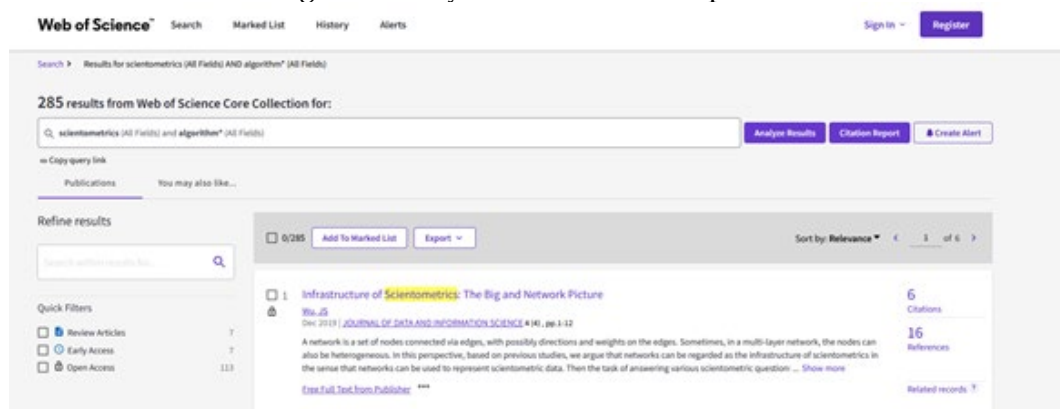


Figure 3c. Results of the search

Subsequently, the data should be downloaded using the option: "Export" and "Plain text file" (Figure 4a). The user then fills the spaces of records. The critical step is to write the numbers of the records; otherwise, the software will save only the first page of results (Figure 3b). Then, the user should select "full record and cited references" (Figure 4b).

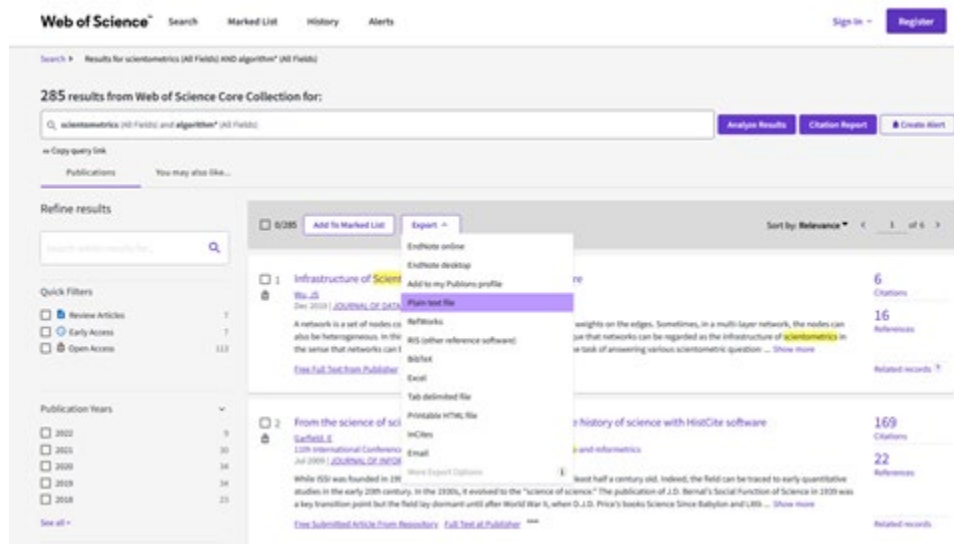


Figure 4a. Save the results format

The critical step in Figure 4a is to select "Save to other file formats."

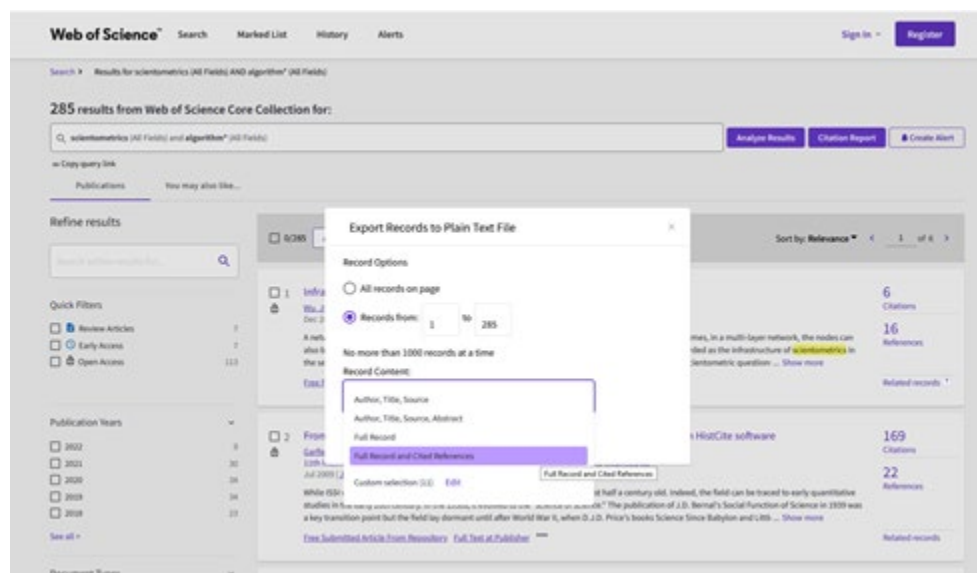


Figure 4b. Export records menu

Fill the records from 1 to 285 and select "Full record and cited references." Once the options are selected, the user clicks on "Export," and WoS will start the download action. Having the .txt file available on the local computer will take a few minutes. At this point, the user has finished the process in WoS. The next step is to upload the .txt file into ToS.

Tree of Science Data Uploading and Visualization

Once the user accesses the tool, they choose the .txt file or drop it onto the dashed square. The Citation counts is 12,298, which corresponds to the number of total seed references (.txt file) (Figure 5). This number should be more than zero to create the tree. If the citations count is zero, the user should re-download the cited referenced from WoS. Please check Figure 4b to select "Full record and cited references." Then, click the "Continue" button to create the tree.

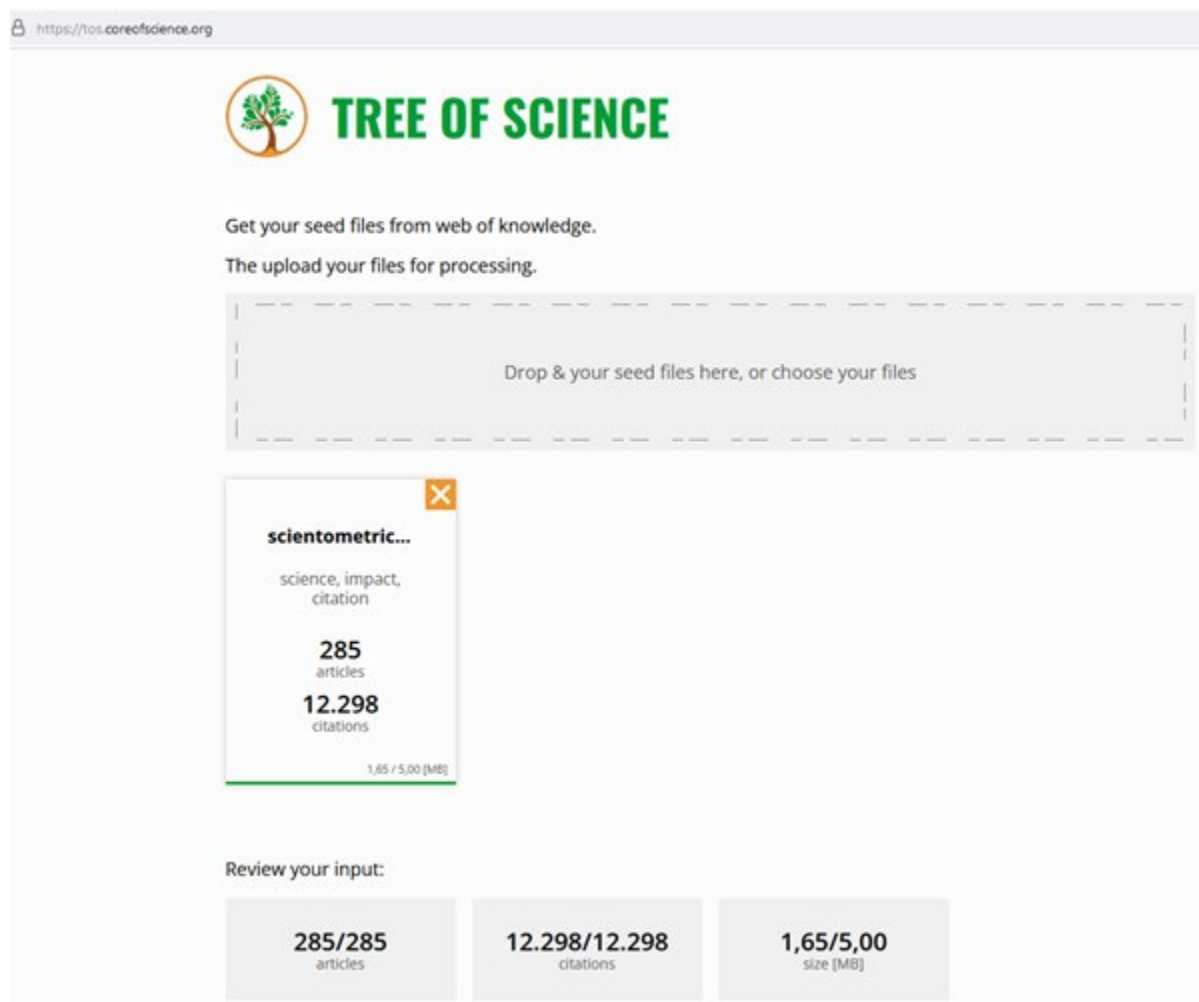


Figure 5. Uploading the results of the search

Finally, the user can start selecting different articles from the “tree” (Figure 6). The DOI is linked so users may access the article text. This example can be found at this link: (<https://tos.coreofscience.org/tree/-MtitUT7NmGCiUnrG1RI>).

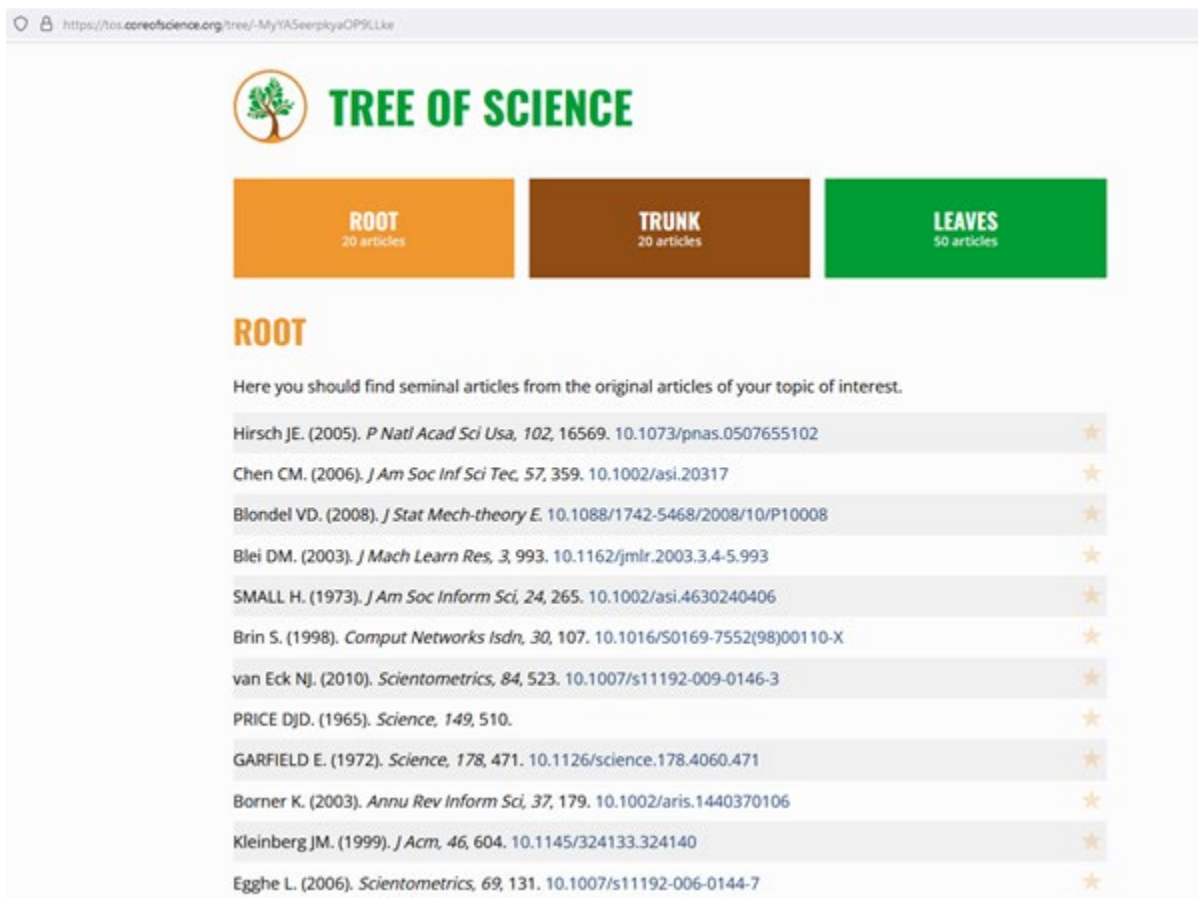


Figure 6. Visualization of the recommended articles (tree table)

Results and Discussion

The principal advantages found in the use of ToS are as follows: First, ToS shows articles despite the database origin (for example, Scopus or Emerald) because the algorithm brings the results from the WoS search, and every single reference cited in each article. Second, time range is not a limitation. Some institutions have restrictions on the time range depending on the contract made with WoS, some of the institutions have access to articles from the last ten years. This is not a limitation for ToS because when the user downloads the file of the search results, it brings all cited references despite the time it was published. Third, ToS presents the results in an organic way (as a tree) that allows intuitive association between each part of the tree and the classification of each article (Classical, structural and recent articles). Finally, ToS is easy to handle. It does not require any specialized skills or applications beyond a web browser. It just needs internet access.

Conclusions

ToS is a valuable tool that users may adopt in order to aid in their evaluation of references for a research design, thesis, or an academic project. ToS can also help early career researchers identify their academic community. ToS helps research studies connect with relevant literature that makes the paper more visible. ToS is an open web-based tool that students, researchers, and academics can use with little training.

The web-based tool has a limitation on the size of the uploaded file, however larger files can be processed using the source code built in R (<https://github.com/coreofscience/r-tos>) or using the R package (<https://cran.r-project.org/web/packages/tosr/index.html>).

References

- Bastian, M., Heymann, S., & Jacomy, M.** (2009). *Gephi: An open source software for exploring and manipulating networks* [Paper presentation]. Third International ICWSM Conference, San Jose, CA, United States. <https://ojs.aaai.org/index.php/ICWSM/article/view/13937/13786>
- Chen, C.** (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377. <https://doi.org/10.1002/asi.20317>
- Duque, P., Meza, O. E., Giraldo, D., & Barreto, K.** (2021). Economía social y economía solidaria: Un análisis bibliométrico y revisión de literatura. *REVESCO. Revista de Estudios Cooperativos*, 138, e75566–e75566. <https://doi.org/10.5209/reve.75566>
- Eggers, F., Risselada, H., Niemand, T., & Robledo, S.** (2022). Referral campaigns for software startups: The impact of network characteristics on product adoption. *Journal of Business Research*, 145, 309–324. <https://doi.org/10.1016/j.jbusres.2022.03.007>
- Gonzalez-Correa, C. A., Tapasco-Tapasco, L. O., & Gomez-Buitrago, P. A.** (2021). A method for a literature search on microbiota and obesity for PhD biomedical research using the Web of Science (WoS) and the Tree of Science (ToS). *Issues in Science and Technology Librarianship*, 99. <https://doi.org/10.29173/istl2679>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A.** (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons. <https://play.google.com/store/books/details?id=cTqyDwAAQBAJ>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D.** (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10), e1–e34. <https://doi.org/10.1016/j.jclinepi.2009.06.006>
- NWB Team.** (2006). *Network workbench tool*. Indiana University, Northeastern University, and University of Michigan.
- Pautasso, M.** (2013). Ten simple rules for writing a literature review. *PLoS Computational Biology*, 9(7), e1003149. <https://doi.org/10.1371/journal.pcbi.1003149>
- Persson, O., Danell, R., & Schneider, J. W.** (2009). How to use Bibexcel for various types of bibliometric analysis. *Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday*, 5, 9–24.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

https://play.google.com/store/books/details?id=ZwZ1_xU3E80C

Porter, A. L., & Cunningham, S. W. (2004). *Tech mining: Exploiting new technologies for competitive advantage*. John Wiley & Sons.

<https://play.google.com/store/books/details?id=-Txp2b7VvAEC>

Robledo, S., Grisales Aguirre, A. M., Hughes, M., & Eggers, F. (2021). “Hasta la vista, baby” – will machine learning terminate human literature reviews in entrepreneurship? *Journal of Small Business Management*, 1–30.

<https://doi.org/10.1080/00472778.2021.1955125>

Sci2 Team. (2009). *Science of science (Sci2) tool*. Indiana University and SciTech Strategies.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>

Valencia-Hernandez, D. S., Robledo, S., Pinilla, R., Duque-Méndez, N. D., & Olivartost, G. (2020). SAP algorithm for citation analysis: An improvement to Tree of Science. *Ingeniería E Investigación*, 40(1), 45–49.

<https://doi.org/10.15446/ing.investig.v40n1.77718>



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).