



There's an App for That

Tree of Science with Scopus: A Shiny Application

Sebastian Robledo

Professor

Universidad Católica Luis Amigó

sebastian.robledogi@amigo.edu.co

Oscar Arbelaez-Echeverri

Researcher

Core of Science

technology@coreofscience.org

Martha Zuluaga

Professor

Universidad Nacional Abierta y a

Distancia

martha.zuluaga@unad.edu.co

Pedro Duque

Professor

Universidad Católica Luis Amigó

pedro.duquehu@amigo.edu.co

Luis Alexander Valencia

Researcher

Core of Science

lavalenciah12@gmail.com

Juan David Alzate-Cardona

Software Engineer

Hourly, Inc.

juanda@hourly.io

Abstract

Tree of Science (ToS) is a scientific literature search tool that produces a small, selected list of citations from a larger pool of citations. Initially developed for searches in the Web of Science, this paper shows how to use it with bibliographic data from Scopus. This new Shiny web application analyzes data from Scopus. It processes a dataset from a Scopus search and creates three reports. The first one shows a descriptive analysis, the second one presents the Tree of Science of the search, and the third one presents a clustering analysis of the three main subtopics. The application is accessible from this link: <https://coreofscience.shinyapps.io/scientometrics/>.

Keywords: Tree of Science, Scientometrics, Scopus

Recommended citation:

Robledo, S., Zuluaga, M., Valencia, L.A., Arbelaez-Echeverri, O., Duque, P., & Alzate-Cardona, J.D. (2022). Tree of Science with Scopus: A shiny application. *Issues in Science and Technology Librarianship*, 100. <https://doi.org/10.29173/istl2698>

Introduction

Researchers and librarians can access millions of research papers. However, processing, selecting, and understanding the content of this data is a difficult and time-consuming task. Therefore, it is essential to use technology to identify the most relevant academic literature. There are several tools, and most of them are split between the point and click interface and code interface. Some examples of software point and click interfaces are CiteSpace ([Chen, 2006](#)), VOSviewer ([van Eck & Waltman, 2010](#)), and SciMAT ([Cobo et al., 2012](#)). However, the most popular programming languages for scientometric analysis are R and Python. Both have specialized packages; for example, R has bibliometrix ([Aria & Cuccurullo, 2017](#)) and litsearchr ([Grames et al., 2019](#)). Examples in Python are ScientoPy ([Ruiz-Rosero et al., 2019](#)) and metaknowledge ([Evans & Foster, 2011](#)).

The ToS algorithm creates a citation network and applies graph metrics to identify papers located in the roots, trunk, and leaves; for a detailed explanation, see Valencia-Hernandez et al. ([2020](#)). ToS has been widely applied in research topics such as entrepreneurship ([Robledo et al., 2021](#)), chemistry ([Durán-Aranguren et al., 2021](#)), management ([Duque et al., 2021](#)), and medicine ([Gonzalez-Correa et al., 2022](#)).

Scopus Search

The first step to creating the ToS of a research topic is searching the Scopus database. Figure 1a presents an example with the word scientometrics. In this case, here are 589 results from the search, see Figure 1b. This number is vital because ToS works best with a number of records between 100 and 600. A minimum number of records (100) is needed to create a citation network; a lower number generates dispersed networks ([Pornprasit et al., 2022](#)). A maximum number of about 600 records is due to the limited memory of Shiny apps (1024 MB); lower specificity will hinder the performance of the algorithm. In the last step, the user must select the BibTeX file, and all the parameters shown in Figure 1c. The “include references” item is key for creating the citation network.

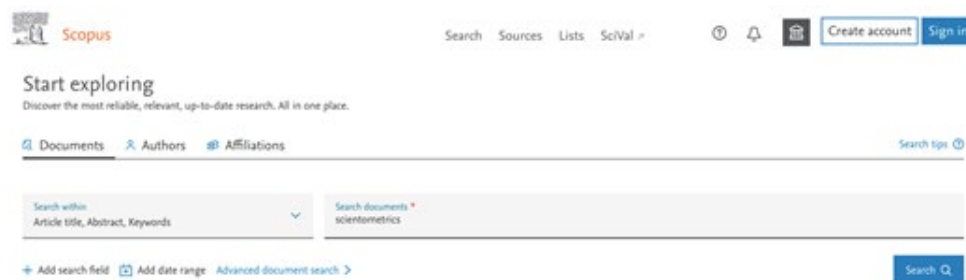


Figure 1a. Example of a search in Scopus database

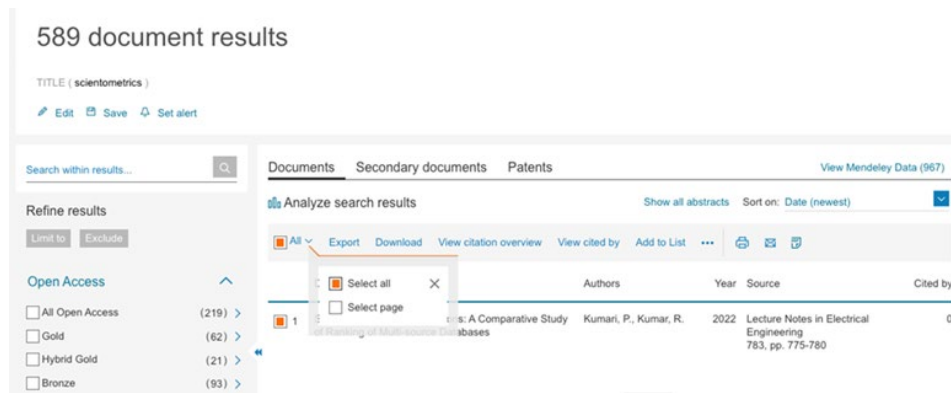


Figure 1b. Selecting the metadata of the papers to be downloaded



Figure 1c. Parameters needed of the data to create the ToS

ToS in a Shiny App

Shiny is an open-source framework to create web apps directly from R ([Chang et al., 2017](#)), and these apps can be uploaded to shinyapps.io to be accessed through a link. Also, shiny developers do not need previous knowledge of JavaScript or HTML to create useful and user-friendly apps. Shiny is used for academics to visualize their research; for professors to teach statistical concepts and big companies in the tech and pharma industry ([Wickham, 2021](#)). Some examples of shiny apps are PeptCreatR ([Arumugaperumal et al., 2022](#)) and DiaThor ([Nicolosi et al., 2022](#)).

Figures 2a-e show the steps for creating the ToS from a Scopus search. Once the user has the BibTeX file from Scopus (the seed of ToS), the user can move forward to the ToS Shiny app following this link <https://coreofscience.shinyapps.io/scientometrics/>. The browse button in Figure 2a opens a new window to upload the BibTeX file. Once the blue bar is completed, Figure 2b, the user can visualize a descriptive analysis in the Importance button, see Figure 2c. This descriptive analysis has the scientific production published each year and the most productive authors and journals. This report is created with the bibliometrix package ([Aria & Cuccurullo, 2017](#)).

The *Evolution - ToS* button presents the papers located in the roots, trunk, and leaves, see Figure 2d. Papers in the roots are seminal, papers in the trunk give structure to the research topic, and papers in the leaves are the current literature. The *link* buttons take the user to a search in Google with the preliminary information from the paper. For example, the seminal papers in scientometrics are Egghe ([2006](#)), Garfield ([1955](#)), and Hirsch ([2005](#)). Egghe ([2006](#)) proposed a new index called g-index to improve the famous h-index proposed by Hirsch ([2005](#)) and Garfield ([1955](#)) was the creator of the Institute of Scientific Information (ISI), nowadays known as Web of Science.

Finally, Figure 2e shows a clustering analysis of the main subtopics. This cluster analysis uses the Blondel et al. (2008) algorithm in the citation network. The Shiny app presents the biggest three clusters (or subtopics) of the seed (research topic) with a word cloud figure to understand the topic of each cluster. The user can change the features of the word cloud, for example, the number of words, their frequency, and remove the unnecessary words.

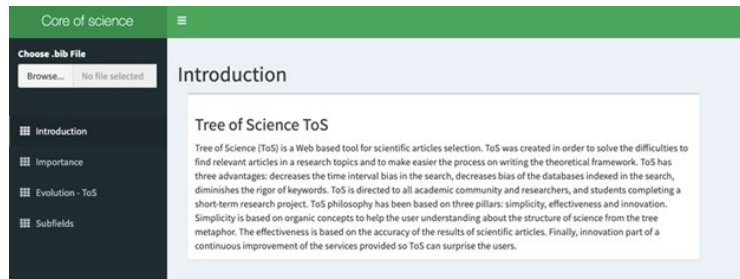


Figure 2a. The landing page of the shiny app

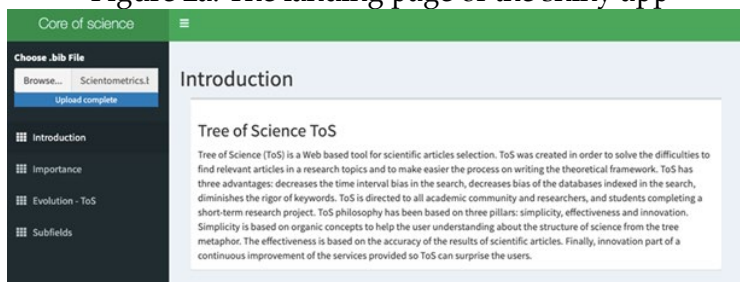


Figure 2b. Seed upload to ToS

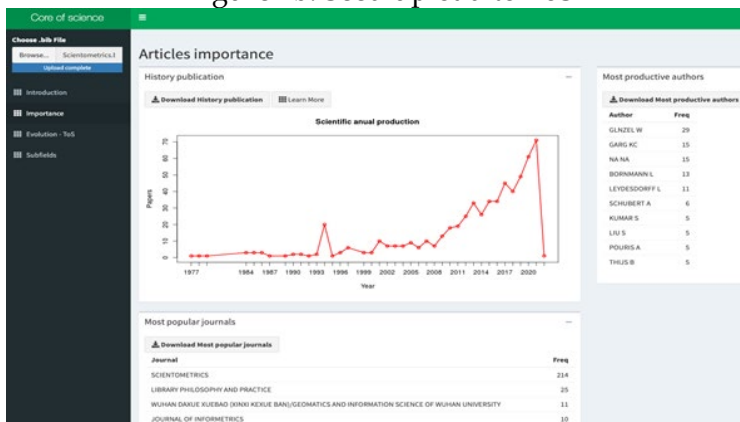


Figure 2c. Descriptive statistics

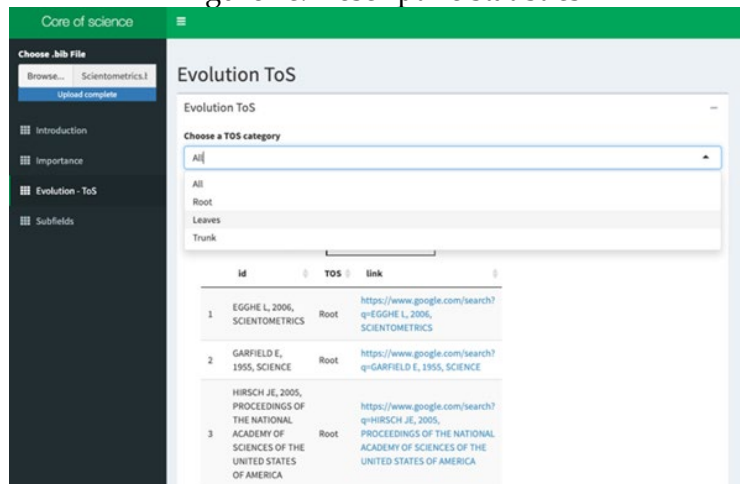


Figure 2d. ToS of the search

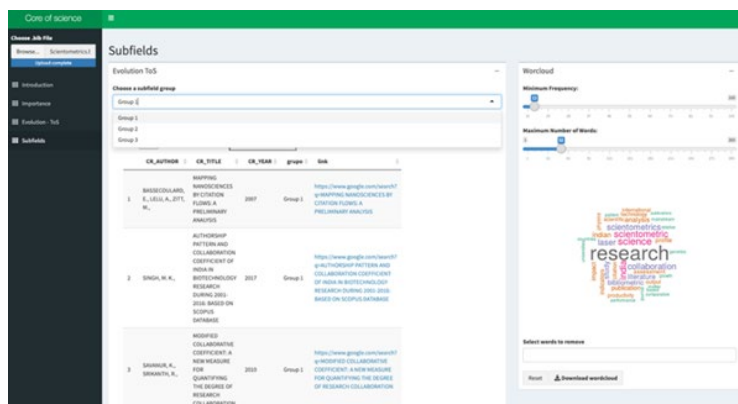


Figure 2e. Cluster analysis

Discussion and Conclusions

ToS was developed as a part of a doctorate thesis, and later the creators decided to start a non-profit organization called Core of Science. The web tool was initially developed with WoS data; however, Scopus is also an important database often available in academic libraries. ToS uses the metaphor of the tree to present the most significant papers from the results in this case obtained from Scopus. Creating a web-based tool is expensive, and most of the time, users must pay this cost. The purpose of the Core of Science is “connecting people through sharing knowledge”; thus, one of the activities is to create free web-based tools for librarians and researchers to help them automate some processes. In this vein, this paper presents a new Shiny app that creates a scientometric analysis to have an overall view of a research topic.

One of the big challenges to creating a citation network with Scopus data is creating a unique identifier of each article and its references. Both should match with other papers in the same search. WoS data has a standard identifier for references, making it more accessible. Also, the references have their DOIs, which facilitates the match among the references and the primary papers.

A limitation of this study is that the ToS algorithm was designed for WoS data, but Scopus data is spread across a broader range of time which implies that some old papers will appear in the trunk because of their publication year. A further improvement of the ToS algorithm could take into consideration this feature in Scopus.

More information about Core of Science is found at: <https://coreofscience.org/>.

References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
<https://doi.org/10.1016/j.joi.2017.08.007>

Arumugaperumal, A., Velayudhan Krishna, D., Alaguponniah, S., Nallaperumal, K., & Sivasubramaniam, S. (2022). PeptCreatR: A web app for unique peptides in human. *International Journal of Peptide Research and Therapeutics*, 28(2), 64.
<https://doi.org/10.1007/s10989-022-10375-4>

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E.** (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J.** (2017). *Shiny: Web application framework for R* (R Package Version 1.5) [Computer software]. R Studio. <https://rdr.io/cran/shiny/>
- 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>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F.** (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63(8), 1609–1630. <https://doi.org/10.1002/asi.22688>
- 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>
- Durán-Aranguren, D. D., Robledo, S., Gomez-Restrepo, E., Arboleda Valencia, J. W., & Tarazona, N. A.** (2021). Scientometric overview of coffee by-products and their applications. *Molecules*, 26(24), 7605. <https://doi.org/10.3390/molecules26247605>
- Egghe, L.** (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152. <https://doi.org/10.1007/s11192-006-0144-7>
- Evans, J. A., & Foster, J. G.** (2011). Metaknowledge. *Science*, 331(6018), 721–725. <https://doi.org/10.1126/science.1201765>
- Garfield, E.** (1955). Citation indexes for science. *Science*, 122(3159), 108–111. <https://www.jstor.org/stable/1749965>
- Gonzalez-Correa, C.-A., Tapasco-Tapasco, L.-O., & Gomez-Buitrago, P.-A.** (2002). 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>
- Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S.** (2019). An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, 10, 1645–1654. <https://doi.org/10.1111/2041-210x.13268>
- Hirsch, J. E.** (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>

Nicolosi Gelis, M. M., Sathicq, M. B., Jupke, J., & Cochero, J. (2022). DiaThor: R package for computing diatom metrics and biotic indices. *Ecological Modelling*, 465, 109859. <https://doi.org/10.1016/j.ecolmodel.2021.109859>

Pornprasit, C., Liu, X., Kiattipadungkul, P., Kertkeidkachorn, N., Kim, K.-S., Noraset, T., Hassan, S.-U., & Tuarob, S. (2022). Enhancing citation recommendation using citation network embedding. *Scientometrics*, 127(1), 233–264. <https://doi.org/10.1007/s11192-021-04196-3>

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>

Ruiz-Rosero, J., Ramirez-Gonzalez, G., & Viveros-Delgado, J. (2019). Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications. *Scientometrics*, 121(2), 1165–1188. <https://doi.org/10.1007/s11192-019-03213-w>

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>

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>

Wickham, H. (2021). *Mastering shiny: Build interactive apps, reports, and dashboards powered by R*. O'Reilly.



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