Short Communications

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)

Carlos Augusto González-Correa
Ph.D. Professor and Senior Researcher of Minciencias
University of Caldas
c.gonzalez@ucaldas.edu.co

Luz-Oleyda Tapasco-Tapasco
Ph.D. Student in Biomedical Sciences
University of Caldas
luz.22916219189@ucaldas.edu.co

Paola Andrea Gómez-Buitrago
Ph.D. Associated Lecturer
University of Cauca
paolagb@unicauca.edu.co

Abstract

In this article, the process and results of a literature search using a new proposed scientific literature search tool (ToS: Tree of Science) aimed at partially overcoming the need to search in different databases was used. In its present form, ToS needs, as input, a previous search in the Web of Science (WoS), and by means of all references used in the articles found in the WoS search, it selects the more significant items, classifying them into three categories: root, trunk and leaves. In our example, from an initial total sum of 164 hits found in WoS, ToS provided 90 items. The following fields of these two results were put together in an Excel sheet for elimination of repetitions and further consideration: title, authors, source, year of publication and DOI (Digital Object Identifier). Then, the titles of the articles were read and graded by the three authors (a senior researcher, a junior researcher, and a PhD student) as 0 (of no interest for the topic), 1 (of possible interest) or 2 (of interest). The marks were added up and those with a score ≥ 3 (56) were selected for the abstracts to be read by the PhD student to

establish a final student's own selection (SoS) of articles for initiating the review of the literature on her topic of interest.

Recommended citation:

Introduction

It is well accepted that good scientific research starts with: a) formulation of a clearly defined question (Lane, 2018), and b) a thorough search, retrieval and review of the scientific literature on the selected topic (Grewal et al., 2016, McKeever et al., 2015, Raich & Skelly, 2013). It is estimated that, so far, humanity has produced millions of scientific papers (estimates vary from at least 50,000,000 up to hundreds of millions, see, for instance Moral-Muñoz et al. (2020)) and that there are more than 26,000 scholarly journals, although some think there may be up to more than double that number (Jinha, 2010). Today, the beginning of the second task is performed via electronic searches (Lu, 2011) in well-known bibliographic databases of scientific literature such as PubMed, Embase, Scopus and WoS for biomedical research. See, for instance, Grewal et al. (2016) and Kraus et al. (2017).

In this article, the authors report on their experience using a new scientific literature analytical tool called the Tree of Science (ToS) (Zuluaga et al., 2016). ToS is based on graph theory, where the vertices, nodes or points are scientific papers, and the edges, arcs or lines are relations between them (Robledo et al., 2014). The output of the system, after doing an appropriate analysis, is a list of articles classified into three categories: root, trunk and leaves, hence the name is ToS. According to the developers, under Root "… you should find seminal articles from the original articles of your topic of interest", under Trunk "…you should find articles where your topic of interest has got a structure, these should be the first authors to discover the applicability of your topic of interest", and, under Leaves "… you should find recent articles and reviews that should condense your topics very well" (Core of Science, 2020). The input for an analysis in ToS is a previous search in the Web of Science (WoS), where too many items are usually retrieved (164 in our case). ToS analyzes all the citations of this search (a total of 13,098 in our case), selects the more relevant ones and classifies them into the three mentioned categories.

Methods

Briefly, the research project of the PhD student involved in the study was to establish if a 6-day colon cleansing protocol (Gonzalez-Correa et al., 2017) produces statistically significant beneficial physiological changes in a group of young, adult, overweight women, and if they are associated with beneficial changes in the intestinal microbiota, expressed as a decrease of the Firmicutes/Bacteroidetes ratio (Koliada et al., 2017). The

first step that we took in our approach was to select terms from the PubMed and EMBASE thesauri (Mesh and EmTree, respectively) and, with them, to build the following search query, run in WoS on 09-21-2021, which produced 164 items:

> KP=((Obesity OR Overweight) AND (Microbiota OR Gastrointestinal Microbiome OR Fecal Microbiota Transplantation OR microflora OR microbiome OR bacterial flora OR feces microflora OR intestine flora OR colon flora OR bacterial microbiome) AND (Human) NOT (bariatric surgery OR bypass OR cancer OR child* OR diabet* OR gestation* OR hormon* OR infant* OR insulin* OR liver OR mice OR mouse OR pig* OR pregnant?* OR rat* OR surg*))

In order to use the results from WoS for a ToS analysis, the following steps have to be followed: Once the search query has been run in WoS, you need to Click on "Export"; then "Plain text File"; Records from 1 to 164 (the number of items retrieved in our example), and finally, "Full Record and Cited References". The platform then downloads a file under the name "savedrecs." This file can be directly uploaded into the ToS platform at https://tos.coreofscience.com/.

From these two lists of articles, the following fields were gathered in a single Excel sheet and duplicate records eliminated: author/s, title, source name, year of publication and DOI (Digital Object Identifier). A subsequent filtering was carried out to eliminate articles that contained words indicating unrelated topics in the title. After this, a list with only the titles was sent to the three authors involved in the project, who separately read them and gave a mark of 0, 1 or 2 to each item, according to the interest that each author assigned to each title in relation to the subject and main aim of the study: 0 = none, 1 = possibly and 2 = relevant and worthy of reading the abstract. To establish the student's own selection (SoS), articles with 3 or more points were considered for further analysis. In the final step, the student read the abstracts and selected those articles that were considered worth reading in full for a review of the topic.

Results

Figure 1 gives a glimpse of the outputs generated by ToS: a) filename, number of articles and number of citations given by the WoS search, once the WoS file is uploaded and recognized by the platform; b) messages shown during data processing after clicking the box "Continue", ("Waiting for a worker to process your tree…", and "Life is hard, we´re getting your data…"), c) banner of the final output giving the number of articles selected and d) list of final results classified into three sections: root, trunk and leaves, showing the first article in each category. (Link to the search: https://tos.coreofscience.com/tree/-MmsziIkoYScFq2vA-7B).
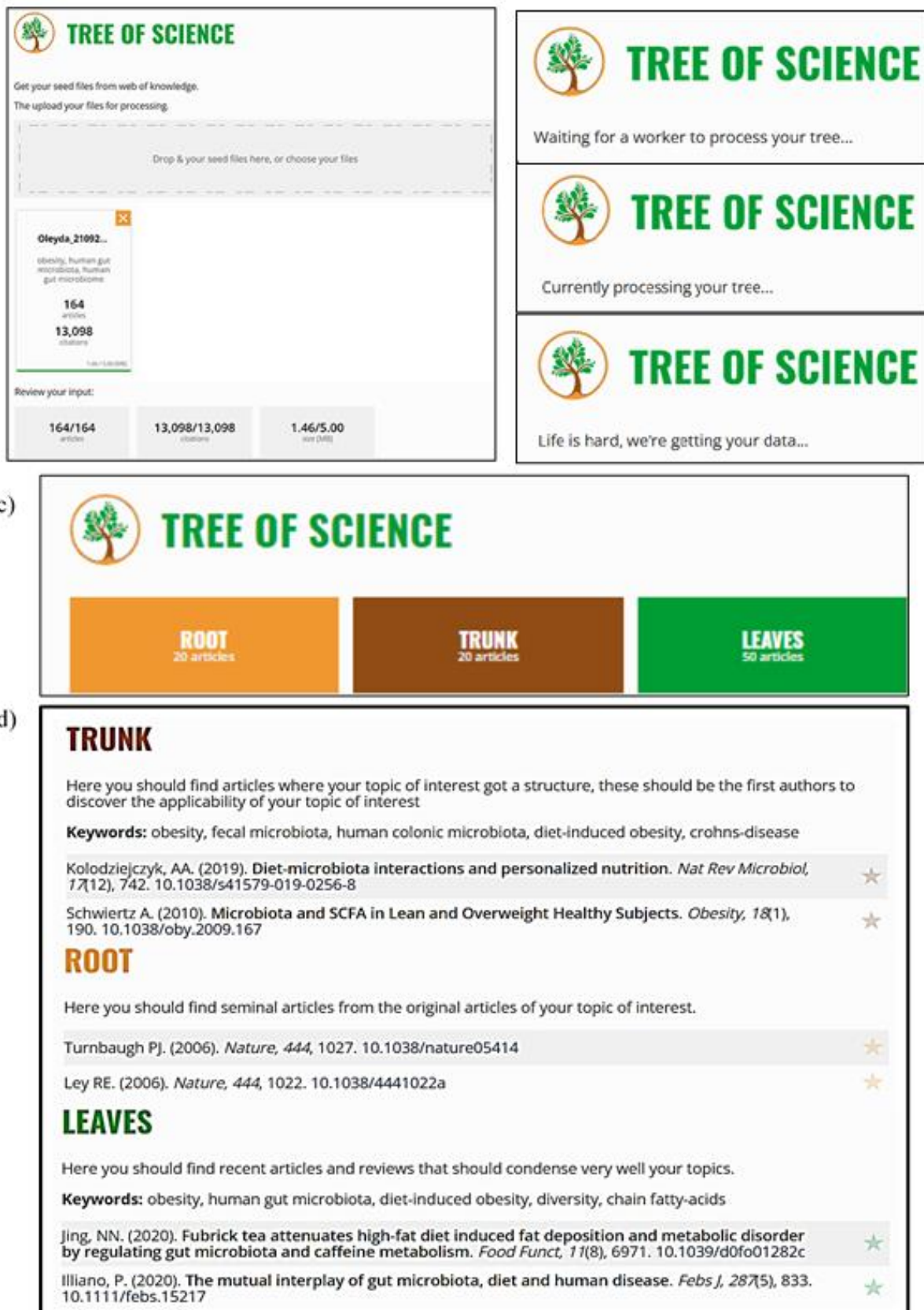
Figure 1: Screenshots after uploading and running the WoS search into ToS.

Figure 2 is a flowchart with the results at the different stages taken from the initial WoS search to the final selection of possible articles of interest for the PhD student's own selection (SoS).

Figure 2. Flowchart from the WoS search to the final result.

START

END

WoS search = 164

-4 without DOI and

For SoS: 56 upper graded (3-6 points) for abstract reading.

ToS search = 90

159 remaining

Number of items per grade:
6 points: 9,
5 points: 15,
4 points: 13,
3 points: 19,
2 points: 12,
1 points: 15,
0 points: 50.

249 – 69 repetitions = 180

-47 items containing at least one of the following words in the title: bariatric (3), bypass /1), cancer (4), child* (7), diabet* (6), dog (3), infant* (2), insulin* (5), liver /3), mice (8), mouse (1), pig* (3), pregnan* (1).

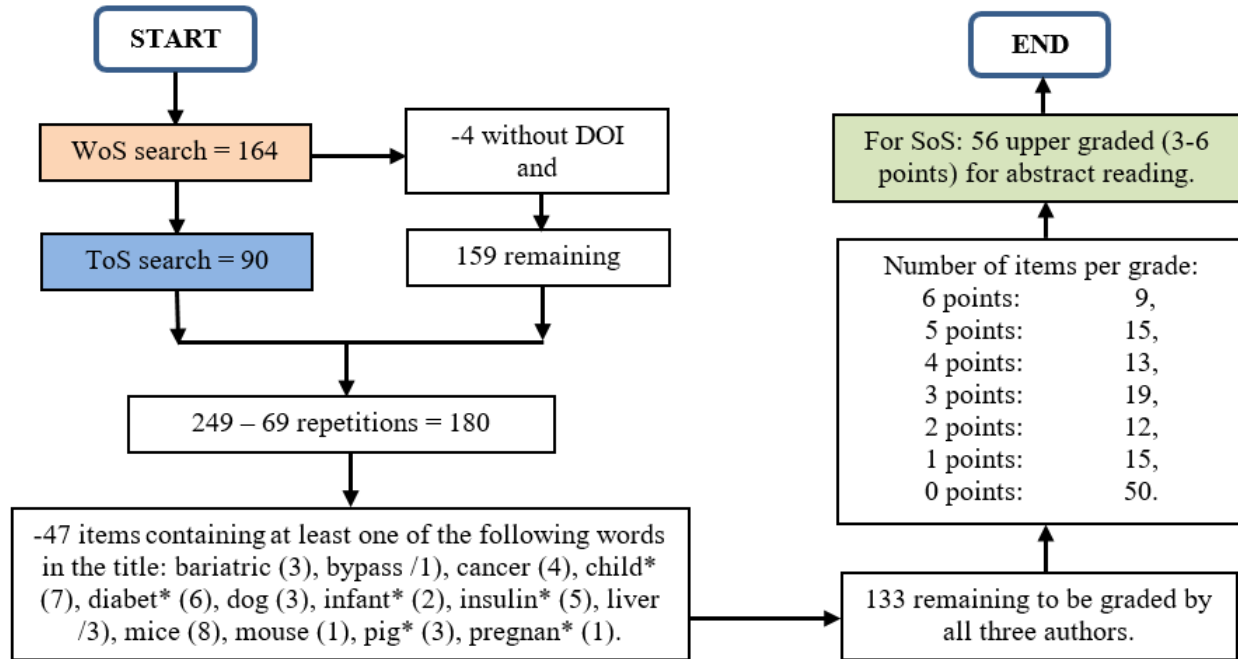133 remaining to be graded by all three authors.

Figure 2: Flowchart from the WoS search to the final result.

Figure 3 is a Venn diagram showing the logical relation between the three data sets obtained in the process: WoS search, ToS search and the final selection of articles to possibly be included in SoS.

Figure 3. Venn diagram of WoS, ToS and SoS results.

SoS
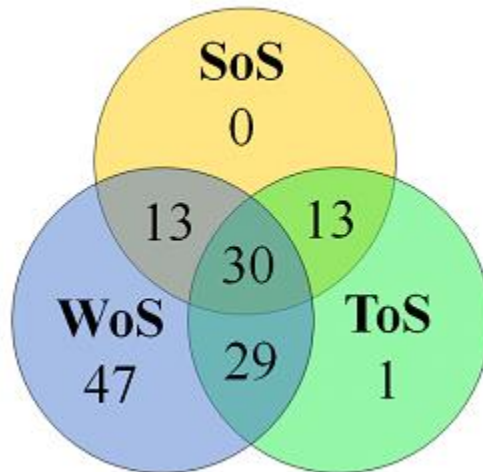0

13   13
30
WoS   ToS
47   29   1

Figure 3: Venn diagram of WoS, ToS and SoS results.

Analysis

Albeit a specific and very stringent search, aimed at excluding unrelated articles from the beginning, the WoS output still gave a relatively large number of items (164). After mixing the results of the WoS and ToS searches, there were still many articles of no

interest for the study´s subject that were manually removed from the list (47). Following the grading of the articles according to the interest assigned by each of the authors, the cutting point for considering articles for further consideration depended on the topic and the specific number given in this step. In this case, deciding to review those with a mark ≥ 3 (56) was agreed, because around 50 articles can be considered an acceptable number for reading their abstracts and deciding which ones merit being read in full. Figure 3 presents the following results: 30 items were both in the WoS and ToS searches that were included in the SoS, 13 items were only present in the WoS list, and 13 were only in the ToS list. Thirteen were only present in either the WoS or ToS list.

Conclusions

Retrieval of pertinent literature for a specific topic is neither an exact science nor an easy task (Grewal et al., 2016, Kraus et al., 2017). Intuition, experience and some creativity are still necessary to obtain what can be considered a good result. This study shows that ToS is an interesting and easy to use tool for the search for scientific literature, partly developed to overcome, to some extent, the need to search individual databases. There are limitations in this exercise as it is a very specific case and the general performance of the different searches will naturally vary according to the subject of interest. Nevertheless, what the authors aimed to achieve in this exercise was to show a way to obtain a good starting point for a beginning PhD student's initial immersion in a topic of interest, and to demonstrate the need for the use of well-structured literature searches when initiating the process of PhD research in the field of biomedical sciences.

References

Core of Science. (2020). *Tree of science*. https://tos.coreofscience.org/

Gonzalez-Correa, C.A., Mulett-Vásquez, E., Miranda, D.A., Gonzalez-Correa, C.H., & Gómez-Buitrago, P.A. (2017). The colon revisited or the key to wellness, health and disease. *Medical Hypotheses*, 108, 133-43. https://doi.org/10.1016/j.mehy.2017.07.032

Grewal, A., Kataria, H., & Dhawan, I. Literature search for research planning and identification of research problem. *Indian Journal of Anaesthesia*, 60(9), 635–9. https://doi.org/10.4103/0019-5049.190618

Jinha, A. Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3), 258–263. https://doi.org/10.1087/20100308

Koliada, A., Syzenko, G., Moseiko, V., Budovska, L., Puchkov, K., Perederiy, V., Gavalko, Y., Dorofeyev, A., Romanenko, M., Tkach, S., Sineok, L., Lushchak, O., & Vaiserman, A. (2017). Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population. *BMC Microbiology*, 17(1), 120. https://doi.org/10.1186/s12866-017-1027-1

Kraus, M., Niedermeier, J., Jankrift, M., Tietböhl, S., Stachewicz, T., Folkerts, H.,

Uflacker, M., & Neves, M. (2017). Olelo: A web application for intuitive exploration of biomedical literature. *Nucleic Acids Research*, 45(W1), W478–83. https://doi.org/10.1093/nar/gkx363

Lane, S. (2018). A good study starts with a clearly defined question: Research question 1 of 2: how to pose a good research question. *BJOG: An International Journal of Obstetrics & Gynaecology*, 125(9), 1057. https://doi.org/10.1111/1471-0528.15196

Lu, Z. (2011). PubMed and beyond: A survey of web tools for searching biomedical literature. *Database*, 2011. 1–13. https://doi.org/10.1093/database/baq036

McKeever, L., Nguyen, V., Peterson, S.J., Gomez-Perez, S., & Braunschweig, C. (2015). Demystifying the search button. *Journal of Parenteral and Enteral Nutrition*, 39(6), 622–635. https://doi.org/10.1177/0148607115593791

Moral-Muñoz, J.A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M.J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, 29(1), e290103. https://doi.org/10.3145/epi.2020.ene.03

Raich, A. & Skelly, A. (2013). Asking the right question: Specifying your study question. *Evidence-Based Spine-Care Journal*, 4(2), 068–071. https://doi.org/10.1055/s-0033-1360454

Robledo, S., Osorio, G.A., & López, C. (2014). Networking en pequeña empresa: Una revisión bibliográfica utilizando la teoria de grafos. *Revista Vinculos*, 11(2), 6–16. https://revistas.udistrital.edu.co/index.php/vinculos/article/view/9664/0

Zuluaga, M., Robledo, S., Osorio, G.A., Yathe, L., Gonzalez, D., & Taborda, G. (2016). Metabolómica y pesticidas: Revisión sistemática de literatura usando teoría de grafos para el análisis de referencias. *Nova*, 14(25), 121. https://doi.org/10.22490/24629448.1735