MAPPING ENVIRONMENTAL SCIENCE: AN OPENALEX REVIEW (Lightning Talk)

Abstract or Résumé:

We reviewed the concepts in OpenAlex under the level 0 Environmental Science category. Using a mixed methods analysis, we graphed the concepts using GraphViz and compared the connections between subcategories with other categorization systems. The problems that leaped out at us were hierarchization, redundancy, and relevance issues, which could make it more challenging to stumble upon relevant information in environmental science research.

1. Introduction

OpenAlex is a free academic bibliographic index with no restriction on metadata retrieval through its API, and it offers a unique machine categorization system for the concepts it has classified (Velez-Estevez et al., 2023). Launched in January 2022, OpenAlex provides access to the corpus from Microsoft Academic Graph that had been available from 2015 to 2021, while continuing to improve the suitability of the services for bibliometric analyses (Scheidsteger & Haunschild, 2023). We examined the categorization of concepts within the Environmental Science level 0 category of OpenAlex and found that while overall the multiple parents structure is useful for an interdisciplinary topic like environmental science, many concepts and categories need to be reviewed for relevance and placement within the hierarchical classification.

2. Methods

Our group investigated the OpenAlex concepts under the Environmental Science level 0 concept to evaluate both the structure of the concepts as a knowledge organization system, and the effectiveness of the concepts as indexing terms for the two hundred and fifty million works indexed by OpenAlex. We used mixed methods, performing some quantitative analysis of the concepts lists and generating graphs of their relationships. We also performed qualitative analysis, examining the graphs, evaluating the suitability of the concept hierarchies, and comparing OpenAlex concepts to other KOSs. Finally, we evaluated the effectiveness of OpenAlex's machine indexing by evaluating the relevance of the indexing terms applied to a sample of articles.

3. Results

Our quantitative analysis found that the concepts are distributed very unevenly among the levels of hierarchy. By graphing the concepts using GraphViz (Fig. 1), we found a variety of problems with the structure of the concepts, ranging from problems with the granularity of the concepts found at the different concept levels to problems with hierarchization and redundancy.

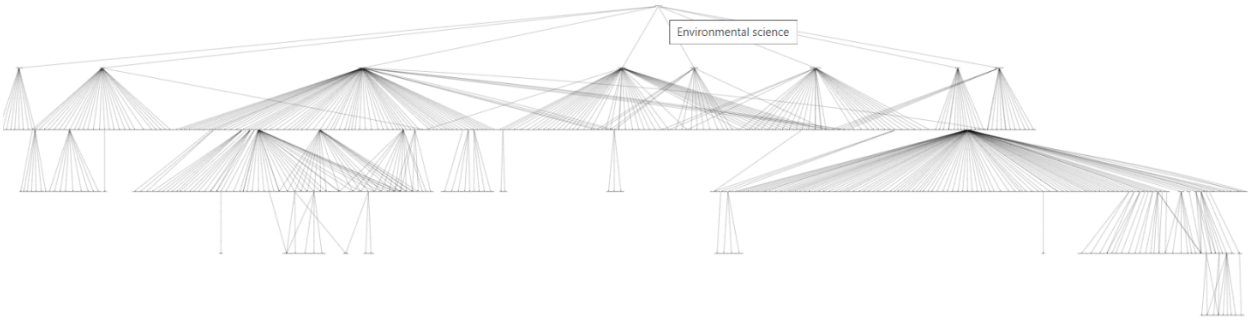

Figure 1: Tree diagram of OpenAlex concepts under Environmental Science

Our qualitative analysis found that the flexibility of OpenAlex concepts having multiple parents made it a good fit for Environmental Science, which is an interdisciplinary field by nature, and restored connections between research in different subdisciplines that might be obscured under more traditional hierarchies. We found that 44% of the semantic relationships between concepts under Environmental Science connected to concepts outside the category of Environmental Science itself. OpenAlex's concept system offers more robust and specific categorization than what is possible in the CRDC, ANZSRC, LoC, or Dewey Decimal systems, although we did find concepts that did not belong in the category or were not effective indexing terms.

Our evaluation of the indexing of articles under Environmental Science found that about half of the indexing terms applied to articles were not correct or appropriate, which would impede the usefulness of OpenAlex searches to researchers looking for articles on a given topic, because of irrelevant search results.

For all these difficulties, the advantages that OpenAlex offers through its interoperability, free and open access, and especially the accessibility of its API, make OpenAlex a valuable tool that merits further development. The machine indexing categories of OpenAlex require some human oversight and adjustment for it to achieve its full potential as a useful tool to researchers.

4. Conclusion

Our mixed methods analysis of the Environmental Science category in OpenAlex found room for improvement in relevance and categorization for many concepts, and revising this classification system could improve upon the open access academic API to increase its usefulness for environmental science research. The way that concepts are classified in the tools that researchers use could shape what information is easily accessible and whether the overall academic conversation about the topic is accurately represented. For example, Korte et al. (2023) found a noticeable shift in OpenAlex from studies describing climate change to research engaging with potential solutions.

The fact that there are flaws in OpenAlex categorization does not necessarily mean they jumped the gun on making it available; a flawed system can still have more value than a system that doesn't exist, and changes can be made to improve it. Before we leap on the OpenAlex bandwagon, we propose improvements to the OpenAlex concepts, which could help researchers find relevant information that prevents them from jumping to conclusions about environmental issues and instead inspires them to leap into action to address them.

References

The Graphviz Authors. (2023). Graphviz. https://graphviz.org/gallery/

Korte, J. W., Bartsch, S., Beckmann, R., Baff, R. E., Hamm, A., & Hecking, T. (2023). From Causes to Consequences, from Chat to Crisis. The Different Climate Changes of Science and Wikipedia. *Environmental Science & Policy*, *148*, 103553-. https://doi.org/10.1016/j.envsci.2023.103553

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. ArXiv. https://arxiv.org/abs/2205.01833

Scheidsteger, T., & Haunschild, R. (2023). Which of the metadata with relevance for bibliometrics are the same and which are different when switching from Microsoft Academic Graph to OpenAlex? *El Profesional de La Informacion*. https://doi.org/10.3145/epi.2023.mar.09

Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, *60*(4), 103385-. https://doi.org/10.1016/j.ipm.2023.103385