Research by the Numbers: Assessing the Performance of Three Products for Citation Analysis

Conte, Marisa; Song, Jean; Townsend, Whitney

http://hdl.handle.net/2027.42/94430
Background

A random sample of 10 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan School of Public Health. The data set included a random sample of 10 names that was selected.

Search Strategies:

- Using "First Name"/"Last Name" search features, enter author's last name
- Select appropriate author from list
- Record total number of publications and h-index from Researcher Profile
- Manually calculate cited counts from citation counts from Researcher Profile

Methodology

A random sample of 10 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan School of Public Health and conducted by a THL librarian. The 10 author names were run against each of the three designed databases: Collexis Research Profiles, Scopus, and Web of Science using the specific search strategies detailed below. Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed below, a further random sample of 10 names was selected.

Limitations

Search strategies were not intended to be comprehensive or capture a definitive list of each author’s research output. Rather, the intent was to utilize search features within each resource for author disambiguation and compare the utility and accuracy of the results.

The different data sources represented in this study each have their own limitations. While none of these limitations represent fatal flaws in the respective products, they are valid concerns or considerations that any end user needs to consider when evaluating the product and its results.

Collexis Research Profiles:

- UMS’s current implementation includes only those authors with a primary appointment to clinical or research units at the University of Michigan Medical School (UMMS).
- While all of the sampled authors met this criteria, the resource is currently not comprehensive for all UMS (UMHS) faculty.
- The current implementation is manually curated to resolve author name disambiguation issues and validated against facutly catalogs to increase publication count accuracy. However, h-index and cited by counts are currently based on ISI Web of Science’s MEDLINE subset, which does not represent full WOS indexing.
- There is no automatic name search within the Research Profiles.

- Cited by counts are not treated and need to be manually calculated.

ISI Web of Science:

- Citation bins using the ‘Distinct Author Sets’ tool are markedly smaller than citation bins using author last name and first initial.
- Many authors do not appear to have been included in the provided author sets.
- Some common names/intial sets are impossible to disambiguate using the ‘Distinct Author Sets’ tool. Full first name and initials are not suggestive, making the disambiguation process for common names even more difficult.

Scopus:

- For name-disambiguation issues, a user can select only up to 15 names. This resulted in the exclusion of many individual publications linked to author names.
- Unclear algorithm for calculating index using different database features. In several instances, a different h-index was obtained from the same set of articles due to an unexplained difference in the number of articles used to calculate h-index.
- A query to Scopus Helpdesk was still open at the time this paper was printed.
- During the period of data collection, the Scopus database presented the following error message: ‘Scopus is currently experiencing issues with the backend and display of some search results. We aim to correct this processing error as soon as possible. Please forgive the inconvenience’ Communication with the Scopus Helpdesk indicated that the problem was fixed before the message was provided throughout the data collection period.

Conclusion

The citation analysis metrics for each of the three tools have unique strengths and weaknesses. Clearly, author name-disambiguation continues to pose a problem, particularly in Scopus and ISI Web of Science. UMS’s implementation of the Collexis Research Profiles goes a long way to addressing this issue through manual curation, but this is a time- and resource-intensive solution that may not be appropriate for all institutions.

Collexis Research Profiles, however, uses only a subset of citations from ISI Web of Science to calculate an h-index and citation metrics. While all three products provide readily available citation metrics clear through their reporting features, end users need to be aware of the limitations of coverage and accuracy when using this information.

Acknowledgements

*Funding provided by the Michigan Institute for Clinical Health Research, NIH Grant U10RR024996.
**Funding provided by the National Center for Integrative Biomedical Informatics, NIH Grant U54MD021159.

Methodology

A random sample of 10 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan School of Public Health and conducted by a THL librarian. The 10 author names were run against each of the three designed databases: Collexis Research Profiles, Scopus, and Web of Science using the specific search strategies detailed below. Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed below, a further random sample of 10 names was selected.

Search Strategies:

- Collexis Research Profiles:
  - Using "First Name"/"Last Name" search features, enter author's last name
  - Select appropriate author from list
  - Record total number of publications and h-index from Researcher Profile
  - Manually calculate cited counts from citation counts from Researcher Profile

- Web of Science:
  - Using "First Name"/"Last Name" search features, enter author's last name and first initial with accents
  - Further disambiguation is necessary, enter first and middle initial with an accent
  - Select relevant papers using the "Select Citations" tool
  - Use the Create Citation Report tool to generate a citation report
  - Record total number of publications, h-index, indexed by citation count

- Scopus:
  - Using Author Search Feature, enter author's last name and first initial
  - Further disambiguation is necessary, enter full name
  - Select relevant journals and select "Citation Tracker" feature
  - Record total number of publications, h-index, indexed by citation count

Selected Results and Discussion Points

- In general, for the sample subset of 10 random researchers used to display the data for this poster:
  - the cited by metric was higher in Scopus than the other two tools
  - the total publications metric was higher in ISI Web of Science than the other two tools
  - variability in one citation metric was an indicator of variability in the other citation metrics

Given that ISI Web of Science is the data source for Collexis' citation metrics, the variability between the h-index and cited by citation counts reflected flaws in the respective products, which will be further investigated.

Research by the Numbers: Assessing the Performance of Three Products for Citation Analysis

Objective

Citation analysis projects constitute a significant portion of the research requests submitted to the Taubman Health Sciences Library (THL). Researchers and administrators are increasingly interested in citation analysis as a way to describe research performance. Citation analysis projects constitute a significant portion of the research requests submitted to the Taubman Health Sciences Library.  The purpose of this study is to compare the citation analysis functions of Scopus and Collexis Research Profiles, in Web of Science, and Scopus in terms of accuracy, coverage, and overall functionality.

Methodology

A random sample of 10 author names was generated from a prior citation analysis request submitted from an administrative unit at the University of Michigan School of Public Health and conducted by a THL librarian. The 10 author names were run against each of the three designed databases: Collexis Research Profiles, ISI Web of Science, and Scopus using the specific search strategies detailed below. Citation analysis metrics for total number of publications, h-index and total cited-by citations were recorded. For the graphs displayed below, a further random sample of 10 names was selected.

Search Strategies:

- Collexis Research Profiles:
  - Using "First Name"/"Last Name" search features, enter author's last name
  - Select appropriate author from list
  - Record total number of publications and h-index from Researcher Profile
  - Manually calculate cited counts from citation counts from Researcher Profile

- ISI Web of Science:
  - Using "First Name"/"Last Name" search features, enter author's last name and first initial with accents
  - Further disambiguation is necessary, enter first and middle initial with an accent
  - Select relevant publications using the "Select Citations" tool
  - Use the Create Citation Report tool to generate a citation report
  - Record total number of publications, h-index, indexed by citation count

- Scopus:
  - Using Author Search Feature, enter author's last name and first initial
  - Further disambiguation is necessary, enter full name
  - Select relevant journals and select "Citation Tracker" feature
  - Record total number of publications, h-index, indexed by citation count

Selected Results and Discussion Points

- In general, for the sample subset of 10 random researchers used to display the data for this poster:
  - the cited by metric was higher in Scopus than the other two tools
  - the total publications metric was higher in ISI Web of Science than the other two tools
  - variability in one citation metric was an indicator of variability in the other citation metrics

Given that ISI Web of Science is the data source for Collexis' citation metrics, the variability between the h-index and cited by citation counts reflected flaws in the respective products, which will be further investigated.

Conclusion

The citation analysis functions for each of the three tools have unique strengths and weaknesses. Clearly, author name-disambiguation continues to pose a problem, particularly in Scopus and ISI Web of Science. UMS’s implementation of the Collexis Research Profiles goes a long way to addressing this issue through manual curation, but this is a time- and resource-intensive solution that may not be appropriate for all institutions. Collexis Research Profiles, however, uses only a subset of citations from ISI Web of Science to calculate an h-index and citation metrics. While all three products provide readily available citation metrics clearly through their reporting features, end users need to be aware of the limitations of coverage and accuracy when using this information.