

Microsoft Academic Graph の書誌情報データとしての評価

文部科学省 科学技術・学術政策研究所 第1研究グループ

塚田尚稔・元橋一之

要旨

本論文では Microsoft 社の書誌情報データ Microsoft Academic Graph (MAG) の利用可能性について、計量書誌学の分野で利用実績の多い Elsevier 社の Scopus をベンチマークとして大規模サンプルで評価した。Open Academic Society から無償ダウンロードできる MAG のバルクデータを用いて、各データベースの論文を DOI で接続して、同一論文 19,166,705 件の書誌情報を比較した。論文出版年は 97.0%、著者数は 98.8%の論文で一致した。参考文献数は Scopus の方が多いが、書誌情報がデータベースに収録されている参考文献に限ると MAG の方が多い。MAG と Scopus のそれぞれから求めた被引用数 (2005 年出版の論文、出版後 10 年間の引用) は、スピアマン順位相関係数が 0.945 であり、高い相関を示す。一方、MAG の 1.66 億件の文献のうちで全ての著者に所属機関情報が存在する論文は 4,373 万 (26.3%) であり、限定的である。MAG は全体として有用なデータベースであるが、現状では、所属機関情報を用いる研究などのためには商用データベースに頼る必要があると考えられる。

Assessment of Microsoft Academic Graph as a Bibliographic Data Source

First Theory-Oriented Research Group, National Institute of Science and Technology Policy (NISTEP), MEXT

ABSTRACT

We assessed Microsoft Academic Graph (MAG) as a bibliographic data source, comparing to Scopus of Elsevier as the benchmark. We used bulk data of MAG, which we can download for free from Open Academic Society. We matched documents in MAG and Scopus using DOI and compared the bibliographic data extracted from MAG and Scopus in terms of 19,166,705 matched documents. As the result, publication years are identical in 97% of the documents. Numbers of authors are so in 98.8% of the documents. Scopus tends to have a larger number of backward reference IDs. But, MAG includes a larger number of bibliographic data of referenced documents. Spearman's rank correlation coefficient between numbers of forward citations calculated from the two databases (as for documents published in 2005, citations in 10 years window) is significantly high (0.945). On the other hand, affiliation data of all authors are included as for only 43.7 million of documents out of 166 million of MAG documents. MAG is very useful database. However, we might need to use proprietary database depending on research objectives.