TCI: Towards a rigorous citation index

Shu-hsien Tseng[1]*, Tzu-heng Chiu [2], Chia-ning Chiang[3]
1 Director General Office, National Central Library
2 Knowledge Services Division, National Central Library
3 Knowledge Services Division, National Central Library
E-mail：yasmine@ncl.edu.tw*

Taiwan Citation Index - Humanities and Social Sciences (TCI) Initiative is a collaborative project of the National Central Library, the Department of Humanities and Social Sciences, Ministry of Science and Technology (DHSS, MOST), and the Science and Technology Policy Research and Information Center (STPI). Citation indexes to scholarly works not only demonstrate the characteristics of literature use in the humanities and social sciences domain, but also establish perspectives rooted in Taiwan's local academic context and Chinese literature.

Building a citation database like the Taiwan Citations Index is very different from conventional bibliographic databases. When building citation indexes, citation entries are found duplicated and variant by nature at the time of data entry. Previous research showed that the earlier years of ISI citation databases still have many unresolved errors and variations. Although Scopus claims better consistency than WoS, researchers also found incomplete editing of overlapping sources for the article references. It is unclear whether they utilize tools to automatically recognize citations or process duplicates. However, it is essential to find a rigorous method to digest the large amount duplicate into a day-to-day process, as well as an effective way to apply the standard record in place of variant citations for the betterment of citation quality.

For accuracy purpose, the TCI team adopted the idea of checking references against the original sources to develop the Search-Apply-Merge module. Consequently, in the TCI system, source records may be used as the benchmark record for citations. This method not only de-duplicates but also improves the accuracy of resulting citation entries. This paper describes the design concept and functions to digest duplicates and obtain better quality on-the-fly for data entry.