Verification of bibliometric methods' applicability for thesaurus construction

Jesper W. Schneider
Department of Information Studies
Royal School of Library and Information Science
jws@db.dk

Available at *http://biblis.db.dk/uhtbin/hyperion.exe/db.jessch04*
356p plus Appendix. ISBM: 87-7415-286-6

The doctoral dissertation work concerns the development and exploration of a semi-automatic thesaurus construction approach based on bibliometric methods.

The main objective of the dissertation is to reintroduce, and further extend, the theoretical and methodological aspects of bibliometric methods to the research area of knowledge organization for the purpose of semi-automatic thesaurus construction.

Thesaurus construction approaches are typically separated into manual approaches and automatic approaches. Albeit, some form of manual thesaurus construction is mandatory due to the relational complexities, semantic ambiguities, and dynamics, inherent in languages. Manual construction and maintenance are complex and time consuming. It is therefore beneficial and necessary to combine manual approaches with automatic approaches due to the complementarity of the two approaches (e.g., Soergel, 1974; Anderson & Pérez-Caraballo, 2001a; 2001b). When automatic approaches are used as a tool for thesaurus constructers, and not as a mean in itself, then we speak of semi-automatic thesaurus construction (Soergel, 1974). This is the foundation for the approach explored in the present dissertation.

In order to pursue the main objective of semi-automatic thesaurus construction, a proposed bibliometric-based methodology of five components is explored as a supplement to manual intellectual thesaurus construction. The methodology is used as a framework for the investigation of the ability of bibliometric methods to identify candidate thesaurus terms and thesaural relationships, as well as to monitor potential terminological and conceptual changes within a specialty area. The bibliometric methods investigated include document co-citation analysis, citation context analysis, co-word analysis, and bibliometric ageing methods. The methodology is explored in a case study of periodontology, a specialty area within dentistry.

The main contributions of the dissertation work are 1) an overall verification of the applicability of co-citation analysis, citation context analysis, and co-word analysis for semi-automatic thesaurus construction; 2) demonstration of the ability of co-citation analysis and citation context analysis to specifically identify important candidate thesaurus terms among a number of potential noun phrases, such terms are important because they are contextual and agreed upon in the scientific community; and 3) demonstration of the ability of co-citation analysis and co-word analysis to detect thesaural relationships

between terms that do not, or rarely, co-occur directly with each other in citation contexts. These results are a direct consequence of the applied bibliometric based methodology.

Consequently, the research reported on in the present doctoral dissertation is a contribution to the development of 'automatic methods' as tools for manual intellectual thesaurus construction.

References

Anderson, J. D. and Pérez-Carballo, J. (2001a). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. Information Processing and Management, 37(2), 231-254.

Anderson, J. D. and Pérez-Carballo, J. (2001b). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: Machined indexing, and the allocation of human versus machine effort. Information Processing and Management, 37(2), 255-277.

Soergel, D. (1974). Indexing languages and thesauri: construction and maintenance. Los Angeles, CA: Melville.