Automatic Classification of Accounting Literature

Abstract: Literature taxonomization is a key element of understanding the knowledge about disciplines. The procedure traditionally used for this classification effort entails a set of manual processes that can be very time consuming and may lead to inconsistent classification. This paper explores the possibilities of using semantic parsing, information retrieval and data mining techniques to develop a methodology for automatic classification of academic articles in accounting based on different criteria. A two-phase experimentation on automatic classification processes has been done in the area of "Treatment", "Accounting Area" and "Mode of Reasoning" (Vasarhelyi et al. 1984, 1988, Brown and Vasarhelyi 1985, Brown et al. 1987). The results from the first phase indicate that using only keywords for classification of accounting literature is not effective. Findings from the second phase indicate that using the full abstract for classification is more successful than using only the keywords. The best results are obtained by using Complement Naïve Bayes (CNB) and Evolving Non-Determinism (END) algorithms which provide accuracy at 81.67%. We discuss the potential path for this preliminary research that seems to be very promising and have several collateral benefits and applications.