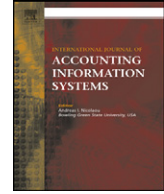




Contents lists available at ScienceDirect

International Journal of Accounting Information Systems



Automatic classification of accounting literature



Vasundhara Chakraborty^a, Victoria Chiu^{b,*}, Miklos Vasarhelyi^c

^a Ramapo College of New Jersey, United States

^b State University of New York at New Paltz, United States

^c Rutgers, The State University of New Jersey, United States

ARTICLE INFO

Article history:

Received 9 November 2012

Received in revised form 19 September 2013

Accepted 4 January 2014

Available online 31 January 2014

Keywords:

Accounting literature

Automatic classification

Taxonomy

Attributes

Semantic parsing

Data mining

ABSTRACT

This paper explores the possibility of using semantic parsing, information retrieval and data mining techniques to automatically classify accounting research. Literature taxonomization plays a critical role in understanding a discipline's knowledge attributes and structure. The traditional research classification is a manual process which is considerably time consuming and may introduce inconsistent classifications by different experts. Aiming at aiding this classification issue, this study conducted three studies to seek the most effective and accurate method to classify accounting publications' attributes. We found results in the third study most rewarding in which the classification accuracy reached 87.27% with decision trees and rule-based algorithms applied. Findings in the first and second studies also provided valuable implications on automatic literature classifications, e.g. abstracts are better measures to use than keywords and balancing under-represented subclasses does not contribute to more accurate classifications. All three studies' results also suggest that expanding article sample size is a key to strengthen automatic classification accuracy. Overall, the potential path of this line of research seems to be very promising and would have several collateral benefits and applications.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The purpose of this study is to develop an automatic classification method to identify the characteristics of accounting and accounting information system research by applying text analytic techniques. Literature taxonomization plays an important role in understanding the knowledge in a discipline; this classification technique assists researchers in examining the development of research areas and disciplines by categorizing

* Corresponding author.

E-mail address: vchiu626@gmail.com (V. Chiu).

documents in several dimensions. It has been used in prior literature to characterize research into specific research subfields in accounting (Birnborg and Shields, 1989; Meyer and Rigsby, 2001; Heck and Jensen, 2007) and summarize the development of major accounting and accounting information system journal publications (Brown et al., 1987; Brown et al., 1989; Vasarhelyi et al., 1988; Previts and Brown, 1993; Badua, 2005; Badua et al., 2011).

The traditional research classification method is a manual process where researchers and/or experts comprehend the article content first and subsequently assign attributes of the taxonomy to the designated manuscript. Categorization of manuscripts is typically based upon a literature taxonomy developed by researchers. The taxonomy enables scholars to arrive at specific characteristics of research in a given field by using it as an academic research index. While the extant research on accounting thought development and evolution applied the traditional manual classification method for long, concerns have been raised in the literature that it is fairly time consuming, costly, and could introduce inconsistent classifications by different researchers (Nobata et al., 1999; Gangolly and Wu, 2000; Krippendorff, 2004; Fisher et al., 2010). Enabling classifications of journal articles automatically could ease these concerns and benefit academic researchers and graduate students in a number of ways. Searches for publications on particular topical areas, research methods and other characteristics of interests could be conducted more efficiently. Summarization and implication of the literature development over time could be examined with enhanced productivity as well. Applying the technique of automatic research classification to accounting and accounting information systems research would support academicians on the identification of literature characteristics and the ability to locate research articles of interest in a more efficient manner.

Acknowledging the criticality of literature taxonomization and the aforementioned research issue, the goal of this study is to refine the process of literature classification and taxonomization to benefit researchers in the accounting information systems and accounting discipline in using classification results. This study was conducted in three phases, to address three main research questions: 1) Using only keywords provided in journal articles, how can the classification process of accounting literature be automated?; 2) Using the abstracts of academic journal articles, how can we automate the process of classification of accounting literature?; 3) Is the accuracy of classification of accounting literature impacted by the combination or choice of elements used in the automation process?

Aiming at seeking the most effective and precise automatic method that classifies accounting research published in multiple accounting and accounting information systems journals, semantic parsing and data mining tools are applied in this study to classify research articles by three taxonomic attribute categories¹ in three studies. Summarizing results from the three phases, we found phase three to be the most promising with the highest degree of automatic classification accuracy of 87.27% by applying decision trees and rule-based algorithms. Findings from the first study suggests that article abstracts provide a better measure of automatic classification compared to keywords. The second study results indicate that applying a balancing approach for under-represented subclasses does not necessarily contribute to a more accurate classification as expected. The third experiment's findings implied that a larger article sample size is very important to improve the accuracy of automatic classification.

Research characterization serves as a valuable information for researchers aiming at revealing the development, trend and evolution of a certain research area or field of knowledge. Through the increasingly populated online and electronic databases, knowledge and data dissemination and communication has been much more prevalent nowadays than decades ago. The benefits of technological advancement on information retrieval and usage, however, still requires more precise exploration and examination by researchers. The contribution of this automation technique on research characteristics classification would extend the usefulness of information retrieval availability by enabling researchers, graduate students and readers to arrive at the numerous characteristics of accounting and accounting information systems research promptly. A widened scope of research and learning needs would be supported; furthermore,

¹ This study compared the automatic classifications with the manual classifications in *Accounting Research Directory (ARD)*, specifically three categories – accounting area, treatment, and mode of reasoning. The ARD (Brown et al., 1994) published accounting literature classifications of 11 accounting and accounting information system journals, see detailed description in methodology section.

Download English Version:

<https://daneshyari.com/en/article/1005421>

Download Persian Version:

<https://daneshyari.com/article/1005421>

[Daneshyari.com](https://daneshyari.com)