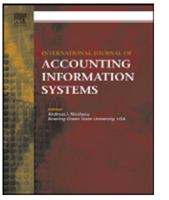
International Journal of Accounting Information Systems 15 (2014) 122–148

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.

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).

1467-0895/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.accinf.2014.01.001

documents in several dimensions. It has been used in prior literature to characterize research into specific research subfields in accounting (Birnberg 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 categories[1]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.

research in the accounting taxonomy and thought development area could specifically benefit from the automatic classification technique by allowing an extensive set of research to be reviewed, categorized and examined with much less constraint due to manual effort limitation.

This study builds upon the literature of both accounting and information science disciplines and sheds light on the approach of sharpening literature classification of accounting and accounting information systems to move towards a more automatic stage. The theoretical ground of this study could be comprehended and linked to the information technology research framework proposed by March and Smith (1995). And to be more specific, this study relates to the methodology "build" and "evaluate" components of the framework as it provides insight on founding an automatic method that is value added to users of academia work. The future path of this line of research seems to be promising and comes with several collateral benefits and applications. The remainder of the study is structured as follows: The next section reviews relevant prior literature and addresses our main research questions. The third section introduces the research methodology and the fourth reports findings from the three studies. Discussion on results summary, limitations and future research implications is provided in the last section.

2. Literature review and research questions

2.1. The traditional classification of accounting literature

Classification of research manuscript characteristics has historically been performed manually to identify specific research characteristics and reveal its overall development scope and trend. The manual classification process requires researchers to comprehend article content first and then assign its representing attributes. It is often necessary for this line of research to involve a heavy process of manual classification on the content of articles by researchers and/or experts and it is fairly time consuming. Research attributes/characteristics development and paradigms evolution have been examined by many accounting researchers through analyzing the article content published in journals in the past decades (Chatfield, 1975; Dyckman and Zeff, 1984; Brown et al., 1987; Vasarhelyi et al., 1988; Brown et al., 1989; Heck and Jensen, 2007; Badua et al., 2011). This section surveys prior literature that applied the traditional classification approach in accounting and relevant business fields, with reviews on its main purpose and used approach as well.

A stream of literature has applied traditional classification on research articles with aims of revealing research attributes and knowledge evolution in publications of specific accounting journals. Chatfield (1975) examined the historical development of research published in the first fifty years of *The Accounting Review (TAR)* and identified four distinct stages of evolution by reviewing publications. Contributions of research published in the *Journal of Accounting Research (JAR)* was analyzed by Dyckman and Zeff (1984); the study concluded that improvement was made mainly on empirical research in accounting, capital markets and refined behavioral research areas in specific between 1963 and 1982. Fleming et al. (2000) examined the evolution of accounting publications by categorizing research methods, financial accounting subtopics, citation analyses, length, and author background of *The Accounting Review (TAR)* journal from 1966 to 1985. Heck and Jensen (2007) is another article that examined TAR's research contribution by taxonomizing publications from 1926 to 2005 by authorship, research methods and accounting topics. Similar content analysis and taxonomization approaches were applied in Brown et al. (1987) to reveal research characteristics and development trend of *Accounting, Organizations and Society (AOS)* from 1976 to 1984. Just et al. (2012) recently examined publications of *Accounting, Organizations and Society (AOS)* from 1990 to 2007 with a combination of analyses on article content classification and other bibliometrics tools such as citation and co-citation analysis.

Other than examining a specific journal, prior literature also provided a comprehensive review of manuscripts by manually classifying publications from multiple outlets. Vasarhelyi et al. (1988) classified accounting literature of 2,136 articles published in six refereed accounting journals by its foundation discipline, school of thought, research methods, and mode of reasoning to systematically examine the historical evolution of research characteristics within 1963–1974. Brown et al. (1989) examined four taxonomies of over 1,100 studies of contemporary accounting literature published in *AOS*, *TAR*, *JAE*, and *JAR* from 1976 to 1984. Literature has also applied the traditional classification method to reveal characteristics of a particular scope of research, school of thought or methodology. Ijiri et al. (1968) surveyed and classified budgeting literature by organization type, budgeting aspects, application areas, and merit evaluation; Felix and Kinney (1982) reviewed auditing literature by a cross-classification approach; auditing literature are categorized by a three-class dimension including state description, model/theory development and hypothesis tests as well as a four-class dimension of auditor's opinion formulation processes. In behavioral accounting research area, Hofstedt (1975, 1976), Birnberg and Shields (1989), and Meyer and Rigsby (2001) classified and examined behavioral studies' research content, research methods and authorship. Besides accounting discipline, finance and information processing fields have applied classification to academic research to help reveal knowledge evolution in sub-fields and over time as well (Ball, 1971; Hakansson, 1973; Gonedes and Dopuch, 1974; Libby and Lewis, 1977; Ashton, 1982; Libby and Lewis, 1982). In summary, classification of literature over time has been a crucial and inevitable technique applied in accounting and information systems research that help shed light on knowledge characteristics and evolution. While extant accounting thought development research has performed manual classification at large, concerns and drawbacks of this approach has been raised, e.g. manual content analysis could introduce researcher subjectivity; high costs of manual processing limits sample size and the power of tests and the generalizability of results (Gangolly and Wu, 2000; Krippendorff, 2004; Fisher et al., 2010). Developing an automatic method for article classification would be an appropriate remedy for these issues. The following section reviews relevant literature in both the information science field and accounting and information systems on the origin, development and application of the automatic classification technique and then the review leads to the three main research questions of this study.

2.2. The automatic classification of terms and documents

The automatic text classification technique is currently still a methodology that has been developing in the literature. Thus far it has not been applied to classify characteristics of academic research in either the accounting/accounting information systems or other disciplines very successfully or extensively.

The development of the automatic approach and its usage on classifying terms and useful thesaurus and generating groups of concepts has been explored in the information science literature decades before its application in the accounting field (Crawford, 1979; Crouch, 1990; Crouch and Yang, 1992; Chen et al., 1995). For example, Crouch (1990) developed a method of automatic thesaurus construction based on the term discrimination value model. Automatic classification method has been proven to produce useful thesaurus class that provides improvement towards information retrieval when it is used to supplement query terms (Crouch, 1990; Crouch and Yang, 1992). Terms and thesaurus have been automatically generated by algorithmic approaches to support the generation and evaluation of systems and network (Chen and Lynch, 1992; Chen et al., 1995). Similar techniques were developed for domain-specific thesaurus for Drosophilia information (Chen et al., 1994) and for computing knowledge base for Worm classification system (Chen and Lynch, 1992). These studies in the information science field have all contributed to the building of the automatic technique in different aspects.

Upon reviewing the literature of accounting and accounting information systems, automatic classification techniques have been found in the application of categorizing financial accounting concepts, conducting textual mining, information retrieval, and readability of financial statements and other accounting information (Gangolly and Wu, 2000; Garnsey, 2006; Garnsey and Fisher, 2008; Fisher et al., 2010). Gangolly and Wu (2000) examined the automatic classification of financial accounting concepts. It provided statistical analysis on term frequencies of financial accounting standards, applied principal components analysis method to decrease dataset dimensionality and used agglomerative nesting algorithm to derive clusters of financial accounting concepts. It is worthwhile to note that the study pointed out that the increasing size of text databases and high cost of domain expertise incurred to develop classifications have necessitated automatic indexing and conceptual classification methods to develop.

Extending the usefulness of automatic classification technique, information retrieval and text mining processes have been used in conjunction with automatic document processing method. Fisher et al. (2010) recently researched on the role of text analytics and information retrieval literature; the study reviewed both manual and computational content analysis of accounting narratives, readability and text-mining studies as well as studies that consists of the extraction of both text elements and quantities imbedded in text from accounting documents. The study implied that computerized content analysis overcomes the limitations introduced by manual content analysis by using software to process electronic textual documents and count word frequency. The relevant literature in the accounting and information systems suggested and implied the need for computerized/automated classification tool for the usage of accounting information, concepts, and documents. However, as in the information science research, “academic literature” have not been used as the main classification object that utilizes automatic classification technique in prior accounting and accounting information systems literature either. And this observed gap in prior research works is where our study provides contribution to.

In line with prior literature that draws inferences from terms in documents, this study explores the automatic classification of research attributes by using research articles' keywords and abstracts. A research article incorporates many subsections; each represents different aspect of the manuscript's research contribution. While the title of any article can create a general impression about the topic of the paper, it does little to answer the content details such as, which foundation discipline the article belongs to or what type of method it applied. Keywords are provided by the authors to enable readers to get an impression on the content of an article, as to which research area the paper belongs to or what techniques were used to conduct the research. Keywords may serve as a dense summary for a document, lead to improved information retrieval, or be the entrance to a document collection (Barki et al., 1993; Hulth, 2003). We expect that using keywords appropriately could facilitate the process of categorization to a great extent. A systematic pattern in the usage of keywords could expedite the process of automatic classification; this argument leads to our first research question:

RQ1: Using only keywords provided in journal articles, how can the classification process of accounting literature be automated?

Academic research articles generally start with an “abstract” which summarizes the salient points of a paper. Prior literature has explored the usage of article abstracts in assisting automatic classification of biology research. Nobata et al. (1999) explored the identification and classification of biology terms by applying information extraction methods that could be generalized to 100 biological abstracts from MEDLINE. The specific techniques adopted to classify terms are statistical and decision tree methods; those used for term candidate identification are shallow parsing, decision trees, and statistical identification methods. The study concluded that applying statistical and decision tree methods to automate biology terms generate the most accurate automatic classification results. They suggested that classification could vary based on the wordlist used as it introduces variance in class types and terms; future studies need to refine algorithms used for automatic classification to reach higher classification accuracy.

Article abstracts can ideally provide crucial information of the research article such as its underlying motivation, research questions, research techniques, main findings as well as implications. Abstracts may serve as convenient resources to generate useful wordlist to categorize article characteristics automatically. The second research question in this study examines this argument:

RQ2: Using the abstracts of academic journal articles, how can we automate the process of classification of accounting literature?

Literature has shown that the extent of accuracy will likely vary based on the different measures applied in automatic classification process (e.g. Nobata et al., 1999; Hulth, 2003). In line with prior studies, we examine the potential automatic classification differences between keyword and abstract results by comparing the overall classification accuracy; the third research question is as follows:

RQ3: Is the accuracy of classification of accounting literature impacted by the combination or choice of elements used in the automation process?

Taken together, the automatic text classification method originated from the information science literature and has been used to aid the process of traditional manual classifications to be more efficient and less inconsistent. While the automatic method has been widely used in many fields, only limited research has examined the automatic research characteristics' classification in accounting and accounting information systems domain. This study aims at contributing to this line of research and expects to improve the classification approach long used in the accounting thought development research and supporting the capacity of the articles that can be classified and examined promptly in research.

3. Methodology

3.1. Sample collection and ARD taxonomy of attributes

Sample articles are hand collected from EbscoHost online academic database and selected based on two main criteria. First, the accessibility of keywords and abstracts in articles from the database; keywords and abstracts of articles are used to generate wordlists to create training dataset in our studies. Second, the availability of research attributes classified in the *Accounting Research Directory (ARD)*. To measure how precise automatic classification of research attributes is, this study uses a set of manual research attributes classification extracted from the *ARD* (Brown et al., 1994)[2] to benchmark against automatic classifications. Aiming at facilitating research efforts of accounting academics and practitioners, the *ARD* provides characteristics of research articles published in twelve leading accounting and accounting information system journals[3] since 1963 by twelve research taxonomies (Appendix I). *ARD* has successfully supported prior literature in understanding characteristics and research contribution of accounting and accounting information systems publications over time (e.g. Brown et al., 1987; Vasarhelyi et al., 1988; Brown et al., 1989; Badua, 2005; Badua et al., 2011).

As a pilot study of automatic research attributes classification in accounting, selecting taxonomic categories that have a fair number of articles represented (that falls under) in each sub taxonomic category is necessary to generate representable results from classification studies. This study examines classifications of three taxonomic categories, *accounting area*, *treatment* and *mode of reasoning*.[4] A sample of 186 articles[5] were used in the first study; the second study applies a balanced class approach encompassing 158 (in accounting area), 453 (in treatment), and 217 (in mode of reasoning) articles. In the final study, a greatly increased sample size was used in an attempt to achieve higher classification accuracy with 763 articles in the accounting area, 627 articles for treatment taxon, and 772 articles in mode of reasoning taxon. Our sample articles are published between 1984 and 2008 and selected from ten accounting journals (Table 1).

3.2. Analyses: a two-phase study of automatic literature classification

Fig. 1 illustrates the two main phases of the study: 1) phase I uses only keywords while and 2) II utilizes the full abstract of journal articles. The first step in the automatic classification process was to develop a parsing tool to extract article keywords and full article abstracts. Parsing is a technique that has been developed in the linguistic and computer science literature to analyze the given text by reasoning out the grammatical structure applied in the text, also known as "syntactic analysis." Semantic parsing allows one to create customized language for specific objectives and is an essential step for Natural Language Processing (NLP), information retrieval, and machine translation (e.g. Tucker, 1984; Trueswell et al., 1994; Salton et al., 1996, and Thompson et al., 1999).

Prior to performing the data mining process some data preprocessing was done.

1) To eliminate unwanted words, a combination of two stop words list was used. The function of a stop word list is to eliminate frequent occurring words that do not have any semantic bearing. The first stop word list applied in the study contains 571 words.[6] The second stop word list is obtained from the Onix

[2] The nature of the ARD's taxonomy and classification carries potential weaknesses that should be taken with caution Manual classification of journal articles has been completed by multiple experts over a span of more than two decades; this may introduce a varied classification precision by experts. That is, the same article may have different designated attributes by different experts. In addition, attributes of research articles that fall under a relatively recent/emerging research area could be identified by ARD taxonomy with challenge. A formalized and automated classification process could standardize the entire procedure.

[3] Journal of Accounting Research, The Accounting Review, Accounting, Organizations and Society, Journal of Accounting Auditing and Finance, Journal of Accounting and Economics, Auditing: A Journal of Practice & Theory, Contemporary Accounting Research, Accounting Historians Journal, Journal of Information Systems, Journal of Accounting and Public Policy, Research in Accounting Regulation, and Journal of Emerging Technologies in Accounting.

[4] Appendix II provides descriptions of three taxonomy categories: accounting area, treatment and mode of reasoning taxons (Brown et al., 1994 and Badua, 2005).

[5] In analysis I, out of the 282 preliminary sample articles selected (based on ARD classified article list), 96 of them were excluded for not providing either keywords or abstracts in the article.

[6] SMART (Salton et al., 1996).

Table 1
Selected accounting journals for articles collection.

Journals	
AOS	Accounting, Organizations and Society
AUD	Auditing: A Journal of Practice and Theory
CAR	Contemporary *Accounting* Research
JAAF	Journal of Accounting, Auditing and Finance
JAE	Journal of Accounting and Economics
JAPP	Journal of Accounting and Public Policy
JAR	Journal of Accounting Research
JETA	Journal of Emerging Technologies in Accounting
JIS	Journal of Information Systems
TAR	The Accounting Review

Text Retrieval Toolkit. Table 2 provides examples of extracted keywords and words from abstracts that are included in the study to classify treatment taxonomic category.

2) After eliminating the stop words, a word count of the remaining words was performed. This step provides the information on how many times a particular word or phrase occurs in an article. If the count for a word is high, it suggests that it is a frequently occurring word. Term frequency was calculated as the next step

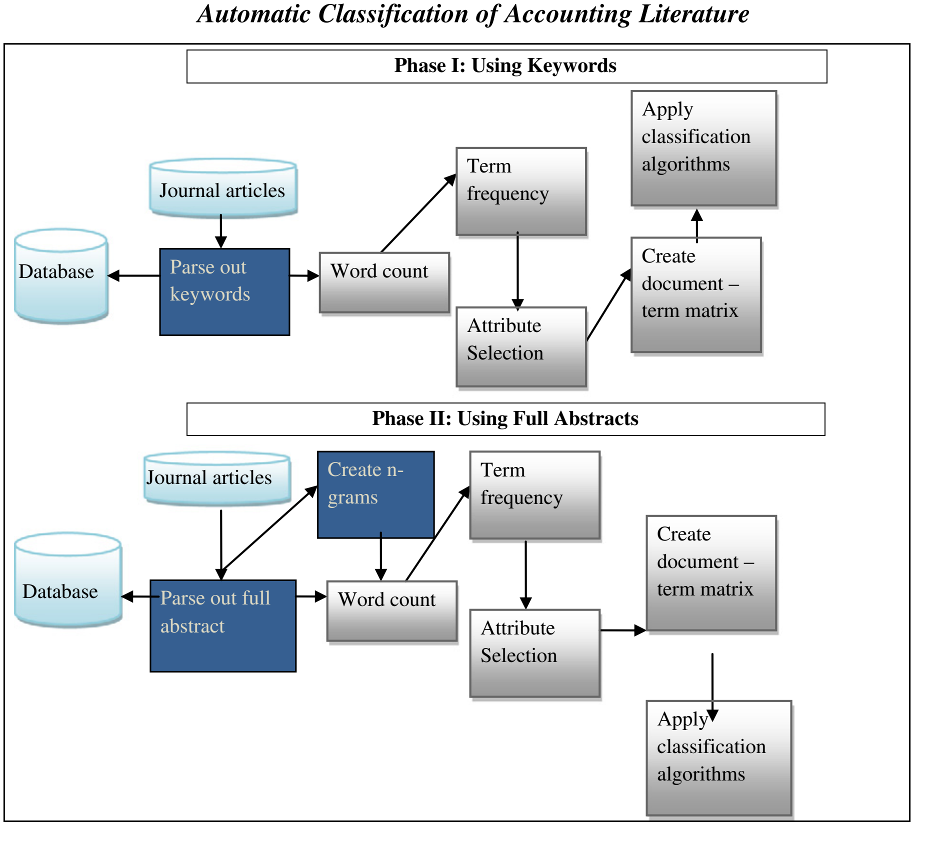


Fig. 1. Two-phase experiment process diagram. The first phase uses keywords and the secsecond phase uses the full abstracts of all collected articles.

Table 2
Examples of keywords used for treatment taxon classification.

Treatment classes	Examples of words from full abstracts
Auditing	Client acceptance, audit partners, hypothesis generation, audit-scope.
Managerial	Decision making, measure, compare, efficiency, evaluation.
Financial	Market reactions, news, speculation, accruals.
Treatment classes	Examples of keywords
Auditing	Auditor independence, risk evalution, risk adaptation, cost of continuous audits, materiality.
Managerial	Decision performance, incentives compensation, mental models, performance measusures.
Financial	Residual income valuation model, economic rents, earnings expectations.

which measures the number of certain word or phrase's occurrences out of the full list of journal articles. This is achieved by adding up the word count from each of the articles for a particular term.[7]

3) Finally, a document-term matrix was created based on term frequency. It is a mathematical matrix which rows correspond to documents and columns correspond to terms; it aims at illustrating the frequency of terms appearing in a collection of documents. An example of a document-term matrix is given in Table 3. As the real document-term matrix created in this study is very large in size, Table 3 provides an example of terms and numbers rather than using the actual data. The first column "File" represents a particular journal article. The column headings from the second column onward show the list of words that occur in various journal articles, e.g. "Residual Income," "Audit-scope," "Materiality" and "Performance measure". Each number in the matrix cell represents how many times a certain word/phrase appears in a particular article, e.g. the word "materiality" occurs two times in the 31.txt document.

In the final step, data mining algorithms were applied to the document-term matrix. Four broad categories of data mining algorithms were used including supervised learning, rule based classifiers, decision trees and other miscellaneous algorithms. A detailed list of the applied algorithms and their corresponding results are provided in Appendix III.

As shown in Fig. 1, the basic steps for Phase I and Phase II are identical. The main difference is that the full abstract is extracted for Phase II whereas only keywords are extracted for Phase I. In case of Phase I there is one additional step after extraction of words from full abstract is the creation of word bi-grams and tri-grams with the objective of finding meaningful phrases that have been used in the articles. The remaining steps in Phase II are essentially the same as in Phase I.

3.3. Validation of automatic classifications

In line with prior literature (e.g. Liu et al., 1998; Nigam et al., 2000; Hall et al., 2009), the validation process of the study was carried out by 1) five-fold cross validation and 2) 66% keywords split for training dataset (with the remaining 37% used for test dataset). The five-fold cross validation first divides the keywords and abstracts into five subsets; the first subset is then used as a test set while the others serve as training set altogether. This process continues by using the second subset as test set and the remaining ones as training set; the iterative process continuous for five times. The 66% split method splits the dataset into two parts — the first part that consists 66% of the data is used for training while the remaining 37% is used for testing (Hall et al., 2009). The next section presents the automatic classifications findings drawn from three studies.

4. Results and analysis

Within the three studies, results obtained from the third study using full abstracts of 282 articles for classification shows the most accurate results. The approach and findings of the three studys follow.

[7] A measure of how often a term is found in a collection of documents. TF is combined with inverse document frequency (IDF) as a means of determining which documents are most relevant to a query. TF is sometimes also used to measure how often a word appears in a specific document.

Table 3
An Example of Document-term Matrix.

File	Residual income	Audit-scope	Materiality	Performance measure
31.txt	6	2	2	1
1400.txt	3	0	2	0
6732.txt	4	0	2	0
8902.txt	0	3	2	1
4569.txt	3	0	0	2
8726.txt	0	4	0	1
7239.txt	7	1	0	0
543.txt	4	1	2	3

4.1. Analysis I: Keywords and abstracts approach

The first study sampled 282 articles with keywords applied in phase I and abstracts applied in phase II of the study (Fig. 1). The algorithms that generated the highest classification accuracy of each taxon are shown in Table 4. The highest classification accuracy was found in *Accounting Area*[8] (85.31%) performed by algorithms that classified the results best was obtained by employing END algorithm and Complement Naïve Bayes of Bayes classifier, both at 81.67%. *Mode of Reasoning* using the OneR algorithm of rule-based classifier group correctly classified the most articles by 56.62% in analysis I of this taxonomic category.[9]

The findings in study I indicate that using abstracts (phase II) to automatically classify articles' attributes results in higher classification accuracy than that of using keywords (phase I). However, an exception occurred in *Mode of Reasoning*; the classification accuracy deteriorated in Phase II in comparison with Phase I which is the opposite of the findings in Accounting Area and Treatment taxons where full abstract (phase II) improves the general results. The least performance in Mode of Reasoning taxon may be explained, to a certain extent, by its many subclasses (11) compared to Treatment (3) and Accounting area (6). These eleven subclasses have a non-uniform data representation. For example, within the 282 articles, only 7 articles used Factor analysis/Probit/Discriminant reasoning method, whereas 158 articles applied Regression analysis method. This non uniform dataset may have prohibited the training and testing datasets in this study to generate results with higher accuracy. To improve the results from analysis I, a probable approach is to expand the study dataset and collect articles with a fairly uniform representation of articles in subclasses of the tested taxons. This may help generate better results, as the training dataset will be more comprehensive after subclasses are populated with more sampled articles and with a more unified representation. The next section elaborates on the second study that applies an expanded dataset with a manipulated balance of classes.

4.2. Analysis II: keywords, abstracts and balanced class manipulation approach

The second study was conducted to improve classification accuracy. The sample size was increased and a balanced class manipulation approach was adopted to tackle the potential negative effect caused by the imbalanced class problem (Tan et al., 2005). The attributes of the sampled accounting literature appears to dominate in a few subclasses of the taxons rather than be evenly distributed among different subclasses. For example, the majority of the sampled articles were classified under "Financial Accounting" in Accounting Area. Under Mode of Reasoning, the dominant methods that were applied in the sampled accounting studies are quantitative methods with a high focus on using "Quantitative: Regression." The results found in Accounting Area, Treatment, and Mode of Reasoning are presented next:

4.2.1. Accounting area

The manipulation affects the articles used in the study involved balancing it across classes to a certain threshold. The ratio of the number of articles between each class to the least popular class in every taxon was calculated. Applying a threshold is necessary prior to the study and therefore this study arbitrarily chose 2.5 as a cutoff point. Table 5a displays the number of articles each class used in the study.

[8] The complete experiment results of the two-phase analysis is in Appendix III.
[9] The complete experiment results of the two-phase analysis is in Appendix III.

Table 4
Analysis I — summary of classification results.

Accounting area		Treatment		Mode of reasoning	
Phase II: Abstracts		Phase II: Abstracts		Phase I:Keywords	
Algorithms	Classification accuracy	Algorithms	Classification accuracy	Algorithms	Classification accuracy
Bayes Classifier-*Complement Naive Bayes*	85.31%	Bayes Classifier-*Complement Naïve Bayes* Miscellaneous-*END*	81.67%	Rule-Based Classifier-*OneR*	56.62%

Results shown in Table 5b suggest that the keywords phase provided better results than abstracts with Naïve Bayes of Bayes classifier providing the highest accuracy at 67.21% under the five-fold cross validation classification method; the second highest results is achieved at 66.67% accuracy by J48 and J48graft algorithms in decision trees classifier group and PART algorithm of rule-based classifier group under the 66% split classification method. The overall findings, however, showed a deterioration of accuracy in comparison to results in analysis (which highest result is at 85.31%).

4.2.2. Treatment

The class distribution of the most dominant to the least popular class in treatment shows a 1.95 ratio which is below the threshold of 2.5, therefore, the original collected dataset (Table 6a) is fully used in the study. Considering that *treatment* is a category that breaks down certain research area to sub-categories, we merged the categories that belong to the same main accounting area into one general class. Specifically, subclasses under treatment taxon numbered #100–#171 were combined to a general financial accounting group, classes numbered #200–#219 altogether represents the general auditing group, and the #300–312 subcategories consist of the general managerial group.

Comparing these results with that of Analysis I, we find substantial reduction in the accuracy of results with the highest level of accuracy being 42%, obtained using random tree and simple cart algorithms of the decision trees family through the use of 66% split classification in keywords phase. The detailed findings are shown in Table 6b.

4.2.3. Mode of reasoning

After manipulation of balancing classes, articles applied in mode of reasoning section totals 218 (Table 7a). The accuracy of results greatly declined from the previous.,the highest accuracy being only 39%, performed by j48 and j48graft algorithms of decision trees classifier group (Table 7b).

4.2.4. Conclusion from studies I & II

Comparing results from the two studies, this study arrived at more favorable results from study I with keywords and abstracts applied to the 282 sampled articles. The highest accuracy level is obtained from the accounting area taxon at 85.31% which is performed by complement naïve Bayes algorithm from Bayes classifier group. While results in mode of reasoning taxon seem less rewarding, findings in treatment

Table 5a
Balancing classes — articles used in each subclass in accounting area taxon.

Accounting area			
Original collected dataset		Manipulated balanced class dataset used	
Tax	32	Tax	32
Financial	288	Financial	38
Managerial	102	Managerial	35
Audit	215	Audit	38
Information systems	15	Information systems	15
Total articles	652	Total articles	158
max/min	19.2	max/min	2.5

Table 5b
Analysis II — summary results of accounting area.

Accounting area				
Algo-rithms/ Results	Five-fold cross validation		66% percentage split	
	Algorithms	Classification accuracy	Algorithms	Classification accuracy
Keywords	Bayes Classifier-*Naïv999e Bayes*	67.21%	Decision Trees Classifiers-*J48 J48graft*	66.67%
	Decision Trees Classifier — *Random Forest*	60.65%	Rule-Based Classifier—*PART*	
	Rule-Based Classifier—*PART*	60.65%	Bayes Classifier—*Naïve Bayes*	42.85%
Abstracts	Bayes classifier—*Bayes Net*	32.60%	Bayes classifier—*Complement Naive Bayes*	38.70%
	Decision trees classifier—*J48 J48graft*	29.34%	Decision trees classifier—*Random Forest*	29.03%
	Rule-Based Classifier—*PART*	28.26%	Rule-Based Classifier—*PART*	22.58%

taxon provide favorable results with Complement Naïve Bayes of Bayes classifier group and END algorithm achieving accuracy at 81.67%.

The second study was conducted for the purpose of improving results from study I, and mainly in the mode of reasoning taxon by balancing the classes in each taxon. However, results have surprisingly deteriorated from study I. One possible reason for this deterioration could be that the limited data availability for study II after applying class-balancing approach. The highest accuracy occurred at the accounting area taxon but only at 67.71%, obtained by the Naïve Bayes algorithm of Bayes classifier family under the five-fold classification method. The less favorable results could possibly be explained by the trade-off between balancing the distribution of classes and the number of articles eventually used in the study. In other words, though the articles dataset has been expanded, the number of articles used in study II post class balance manipulation turns out to be less than that applied in study I.

Hence, it could be said that the assumption that inaccuracy of results was due to the negative effect of imbalanced class issue on classification accuracy seemed less significant than our earlier prediction. The final study aimed at improving results obtained by increasing the sample size used. This increased sample size approach in the last study yielded by far the best results in terms of classification accuracy.

4.3. Final study — increased sample with abstracts applied

The final study was built upon the first study, by using a sample size of articles three times more than study I. Specifically, 763 articles were used for accounting area taxon, 627 articles for treatment taxon, and 772 for mode of reasoning. Since studies showed that phase II (using abstracts) provided better results than phase I (using keywords), this study was conducted solely by using abstracts retrieved from the collected articles to provide automatic classification results.

4.3.1. Accounting area

The findings slightly improved from the results obtained in analysis I. Complement Naïve Bayes Algorithm of Bayes classifier once again provides the best result with accuracy of 85.78% under the 66% split method. The second and third highest results of accuracy are performed by the J48graft of decision tree

Table 6a
Balancing classes — articles used in each subclass in treatment taxon.

Treatment			
Original collected dataset		Manipulated balanced class dataset used	
#100–#171 General Financial Accounting	191	#100–#171 General Financial Accounting	191
#200–#219 General Auditing	164	#200–#219 General Auditing	164
#300–#312 General Managerial	98	#300–#312 General Managerial	98
Total	453	Total	453
max/min ratio	1.95	max/min ratio	1.95

Table 6b
Analysis II — summary results of treatment.

Treatment				
Algorithms/ Results	Five-fold cross validation		66% percentage split	
	Algorithms	Classification accuracy (%)	Algorithms	Classification accuracy (%)
Keywords	Decision Trees Classifier—*Random Forest*	41.68	Decision Trees Classifier—*Random Tree Simple Cart*	42.00
	Bayes Classifier—*Naïve Bayes NaïveBayes Updateable*	38.37	Bayes Classifier—*Naïve Bayes Naïve Bayes Multinomial Naïve Bayes Multinomial Updateable Naïve Bayes Updateable*	40.90
	Rule-Based Classifier—*JRip*	37.20	Rule-Based Classifier—*JRip*	
Abstracts	Bayes Classifier—*Complement Naive Bayes Naïve Bayes Multinomial*	30.63	Bayes Classifier—*Complement Naive Bayes* Rule-Based Classifier—*OneR*	26.11
	Decision Trees Classifier—*Random Forest*	23.03		
	Rule-Based Classifier—*JRip*	29.62	Decision Trees Classifier—*Random Forest*	20.89

classifier group at 81.42% and followed by Bayes Net algorithm of Bayes classifier at 80/96, both used five-fold cross validation method of classification (Table 8).

4.3.2. Treatment

Under treatment taxon, the most accurate classification result has been achieved in the final study as well. Both J48graft algorithm of decision trees classifier and Ridor algorithms of rule-based classifier accurately classified 87.27% of articles under the five-fold cross validation method.

The second best result was reached by using the DMNB Text algorithm of Bayes classifier group with 86.67% of articles correctly classified. J48graft algorithm of decision tree family is the third best algorithm used which also ranks the highest among the 66% split section by accurately classifying 85.71% of articles. J48graft is followed by DMNB Text and Ridor algorithm which provide accurate results at 83.93% (Table 9).

4.3.3. Mode of reasoning

The level of accuracy improved most significantly under mode of reasoning taxon in the final study in comparison with the prior two studies (Table 10). The highest accuracy level at 70.18% is obtained from both the Naïve Bayes Multinomial and Naïve Bayes Multinomial Updateable of Bayes classifier group under the five-fold cross validation classification method. This finding has improved nearly 14% from the result shown in the first study (56.62%).

Table 7a
Balancing classes — articles used in each subclass in mode of reasoning taxon.

Mode of reasoning			
Original collected dataset		Manipulated balanced class dataset used	
Quantitative: DS	51	Quantitative: DS	33
Quantitative: Regression	270	Quantitative: Regression	35
Quantitative: Anova	73	Quantitative: Anova	33
Quantitative:Factor Analysis/MDS/ Probit/Discriminant	19	Quantitative:Factor Analysis/MDS/Probit/ Discriminant	19
Quantitative: Non-Parametric	16	Quantitative: Non-Parametric	16
Quantitative: Correlation	14	Quantitative: Correlation	14
Quantitative: Analytical	107	Quantitative: Analytical	35
Qualitative	49	Qualitative	33
Total	599	Total	218
max/min ratio	19.28	max/min ratio	2.5

Table 7b
Analysis II — summary results of mode of reasoning.

Mode of reasoning				
Algorithms/ results	Five-fold cross validation		66% percentage split	
	Algorithms	Classification accuracy (%)	Algorithms	Classification accuracy (%)
Keyword	**Decision Trees Classifier—*J48 J48graft***	**39.00**	**Bayes Classifier—*Naïve Bayes***	**35.29**
	Bayes Classifier—*Bayes Net*	38.89	Decision Trees Classifier—*BF Tree*	29.41
	Rule-Based Classifier-*PART*	36.00	Rule-Based Classifier—*Decision Table-Ridor*	
Abstract	**Decision Trees Classifier**—*J48 J48graft*	**28.24**	**Bayes Classifier**—*DMNB Text*	**31.11**
	Bayes Classifier—*Naïve Bayes Naïve Bayes Multinomial Naïve Bayes Multinomial Updateable Naïve Bayes Updateable*	27.48	Rule-Based Classifier—*PART*	22.22
	Rule-Based Classifier—*PART*	25.95	Decision Trees Classifier—*Random Forest*	20.00

The second highest result is obtained by applying the 66% split method, a 65.45% of accuracy was performed by the DMNB algorithm of Bayes classifier group which is followed by the random forest algorithm (decision tree group) that performed the third best at 64.22%.

5. Conclusion and implications

This article develops a literature classification technique to identify and categorize characteristics of accounting and accounting information systems research automatically. It uses semantic parsing and data mining techniques to explore the possibilities of developing a methodology to automatically classify academic articles in accounting, on the basis of various criteria and taxons. Three studies were conducted to examine and refine the classification process. Keywords and abstracts of research articles were both applied to the first two studies while in the final study only abstracts were used.

A summary of findings in the three studies are illustrated in Table 11. The final study provides most accurate results with 87.27% of sampled articles correctly classified by J48graft (decision trees) and Ridor (rule-based) algorithms in the treatment taxon under five-fold cross validation. Complement Naïve Bayes of Bayes classifier performed very well in accounting area classification by reaching accuracy of 75.78% under the 66% split approach. In mode of reasoning taxon, Bayes algorithms have provided results that improved greatly from remaining studies, which achieved a correct classification level of 70.18%. The overall results obtained in the final study outperformed the first two significantly, suggesting that using abstracts is more effective than keywords in generating the training set to automate the classification process. In addition, results suggest that expanding sample size is a critical issue that contributes to a better automatic classification performance.

Taxonomization of literature has been critical in many disciplines. However, prior literature classified research works manually, which is a process that is fairly time consuming and could lead to classification inconsistencies. The findings in this study seem promising and indicate that the aforementioned limitations can be improved by automatically classifying the literature. This study is a contribution to the general accounting literature and accounting information systems literature in particular. The analyses conducted

Table 8
Final study—summary results of accounting area.

Accounting area			
Five-fold cross validation		66% percentage split	
Algorithms	Classification accuracy (%)	Algorithms	Classification accuracy (%)
Decision-Trees Classifier—*J48graft*	81.42	Bayes-Classifier—*ComplementNaive Bayes*	85.78
Bayes Classifier—*Bayes Net*	80.96	Decision-Trees Classifier—*J48graft*	79.05
Rule-Based Classifier—*Ridor*	77.98	Rule Based Classifier—*Ridor*	72.97

Table 9
Final study — summary results of treatment.

Treatment			
Five-fold cross validation		66% percentage split	
Algorithms	Classification accuracy (%)	Algorithms	Classification accuracy (%)
Decision-Trees Classifier—*J48graft*	87.27	Decision-Trees Classifier—*J48graft*	85.71
Rule-Based Classifer—*Ridor*		Bayes Classifier—*DMNB Text*	83.93
Bayes Classifier-*DMNB Text*	86.67	Rule Based Classifier—*Ridor*	

use semantic parsing and classification methodologies, and the results carry impact on the broader accounting literature by facilitating and strengthening the process of research classification. A broader range of learning and research needs could be satisfied through the application of this automatic literature classification technique; especially for research conducted in the accounting taxonomy and thought development topical areas, for example, it allows researchers to examine and review much greater volume of literature with minimum manual effort involved. Analysis would be supported by the many available online electronic databases in this age. Furthermore, findings not only reveal the usefulness of semantic parsing and data mining tools on refining accounting literature classification, which enable researchers, graduate students and readers to promptly arrive at the numerous characteristics of accounting and accounting information systems research, but also could lead to implications on potential emerging taxonomic classes suggesting future research development directions.

Future research can continue to build on this study by exploring the automation of classification with other criteria or taxons of accounting literature applied. Additionally, developing techniques with higher precision, benefiting other disciplines by applying automatic taxonomization of publications, sharpening the tools for analyzing research evolution etc., are other directions in which further research can make contribution on. An integral part of developing similar classification processes for different taxons, with several subclasses, would be building a training dataset that has a uniform representation of the different classes. It would also be interesting to explore whether any new areas of research are developing, meaning that whether new classes/taxons need to be added to the taxonomy. Developing an automated method to explore this type of research would be an interesting extension of this study as well.

One of the limitations of this study is the insufficient number of articles sampled in developing the classification methodology. As demonstrated by the results from studies, classification accuracy improves significantly with the increase of the number of articles in the data corpus. Thus, downloading and sampling an even larger set of articles would likely to increase classification accuracy more.

Phase II with the adoption of article abstracts leads to better results in the end compared to using keywords. However, abstracts generally have restrictions on the number of words and as a result may not contain sufficient information to accurately classify articles' attributes. With the inclusion of other sections of an article such as the conclusion, results, methodology or even the full article for classification may likely have a positive impact on the results. Future research could consider these factors in the study.

Results from Phase I suggest that the use of keywords is far less effective in correctly classifying journal articles as compared to article abstracts. While keywords have been taken as measures of research articles'

Table 10
Final study — summary results of mode of reasoning.

Mode of reasoning			
Five-fold cross validation		66% percentage split	
Algorithms	Classification accuracy	Algorithms	Classification accuracy
Bayes Classifier—*Naïve Bayes Multinomial*	70.18%	Bayes Classifier—*DMNB Text*	65.45%
Naïve Bayes Multinomial Updateable			
Decision Trees Classifier—*Random Forest*	64.22%	Decision Trees Classifier—*Random Forest*	62.83%
Rule-Based Classifier—*ZeroR Conjunctive Rule*	63.99%	Rule-Based Classifier—*JRip*	

Table 11
Studies results comparison.

Results comparison	Accounting area		Treatment		Mode of reasoning	
	Algorithms	Classification accuracy	Algorithms	Classification accuracy	Algorithms	Classification accuracy
Final study (abstracts)	66% percentage split Bayes Classifier—*Complement Naive Bayes*	85.78%	Five-fold cross validation Decision Trees Classifier—*J48graft* Rule-Based Classifer—*Ridor*	87.27%	Five-fold cross validation Bayes Classifier—*Naïve Bayes Multinomial Naïve Bayes Multinomial Updateable*	70.18%
Analysis I	Phase II: Abstracts/66% Split Bayes Classifier—*Complement Naive Bayes*	85.31%	Phase II: Abstracts/66% Split Bayes Classifier—*Complement Naïve Bayes* Miscellaneous-*END*	81.67%	Phase I:Keywords/66%Split Rule-Based Classifier—*OneR*	56.62%
Analysis II	Phase I: Keywords/66% Split Decision Trees Classifiers—*J48 J48graft* Rule-Based Classifier—*PART*	66.67%	Phase I: Keywords/66% Split Decision Trees Classifier—*Random Tree Simple Cart*	42.00%	Phase I: Keywords/Five-fold Decision Trees Classifier—*J48 J48graft*	39.00%

content in the literature (Hulth, 2003), the finding in this study is a pointer towards the inadequacy of keywords being used to describe articles. Future research needs to further assess how and what kind of keywords should be used to correctly represent article content. Creating a set of keywords and surveying authors to find out their opinion about using them in their articles or suggesting the list to be used by academic databases are some of the measures that could be incorporated in future research.

Finally, there may be some judgmental differences between different experts that have coded the articles used in this study over a prolonged period of time. As a result, different experts could have categorized similar articles under different taxons and this could have had a bearing on the final results in this study. The accuracy of automatic classification was validated against the manual codes provided by the experts and if these manual codes have judgmental differences then it would in turn affect the results of automatic classification. To fine tune and arrive at a robust standardized method for automatic article classification, future research should consider using multiple domain experts to reclassify all the journal articles used in the training dataset to assure that judgmental errors do not bias the results.

Appendix I

I.A. Taxonomy Classes

A. RESEARCH METHOD

1. Analytical Internal Logic
2. Analytical Simulation
3. Archival — Primary
4. Archival Secondary
5. Empirical Case
6. Empirical Field
7. Empirical Lab
8. Opinion Survey
9. Mixed

B. Inference Style

1. Inductive
2. Deductive
3. Both

C. Mode of Reasoning

1. Quantitative: Descriptive Statistics
2. Quantitative: Regression
3. Quantitative: Anova
4. Quantitative: Factor Analysis, MDS, Probit, Discriminant
5. Quantitative: Markov
6. Quantitative: Non-Parametric
7. Quantitative: Correlation
8. Quantitative: Analytical
10. Mixed
90. Qualitative

D. Mode of Analysis

1. Normative
2. Descriptive
3. Mixed

E. School of Thought

1. Behavioral — Hips
2. Behavioral — Other
3. Statistical Modeling — EMH
4. Statistical Modeling Time Series
5. Statistical Modeling Information Economics
6. Statistical Modeling Mathematical Programming
7. Statistical Modeling Other
8. Accounting Theory
9. Accounting History
10. Institutional
11. Other
12. Agency Theory
13. Expert Systems

F. Information

100. Financial Statements
101. Net Income or EPS
102. Income Statement
103. Balance Sheet
104. Cash Flows, Etc.
105. Other Fin. Statement
106. Financial Ratios
107. Combinations 1–2
108. Quarterly Reports
109. Foreign Currency
110. Pension
112. Debt Covenants
200. Internal Information
201. Performance Measures
202. Personality Measures
203. Auditor Behavior
204. Manager Behavior
205. Decision Making
206. Internal Controls
207. Costs

208. Budgets
209 Group Behavior
210. Pricing
211. Compensation
300. External Information
301. Footnotes
302. Sec Info (10 K)
303. Forecasts
304. Audit Opinion
305. Bond Rating
309. Other
400. Market Based Info
401. Risk
402. Security Prices or Return
403. Security Trading
404. Options
405. All Of The Above-Market
500. Mixed

G. Treatment
100. Financial Accounting Methods
101. Cash
102. Inventory
103. Other Current Assets
104. Property, Plant and Equipment/Depreciation
105. Other Non-Current Assets
106. Leases
107. Lon6 Tern Debt
108. Taxes
109. Other Liabilities
121. Valuation (Inflation)
122. Special Items
131. Revenue Recognition
132. Accounting Changes
133. Business Combinations
134. Interim Reporting
135. Amortization/Depletion
136. Segment Reports
137. Foreign Currency
141. Dividends-Cash
143. Pension (Funds)
150. Other — Financial Accounting
160. Financial Statement Timing
170. R & D
171. Oil & Gas
200. Auditing
201. Opinion
202. Sampling
203. Liability
204. Risk
205. Independence
206. Analytical Review
207. Internal Control
208. Timing

209. Materiality
210. EDP Auditing
211. Organization
212. Internal Audit
213. Errors
214. Trail
215. Judgement
216. Planning
217. Efficiency — Operational
218. Audit Theory
219. Confirmations
300. Managerial
301. Transfer Pricing
302. Breakeven
303. Budgeting & Planning
304. Relevant Costs
305. Responsibility Accounting
306. Cost Allocations
307. Capital Budgeting
308. Tax (Tax Planning)
309. Overhead Allocations
310. HRA/Social Accounting
311. Variances
312. Executive Compensation
400. Other
401. Submissions To The FASB Etc.
402. Manager Decision Characteristics
403. Information Structures (Disclosure)
404. Auditor Training
405. Insider Trading Rules
406. Probability Elicitation
407. International Differences
408. Form Of Organization (Partnership)
409. Auditor Behavior
410. Methodology
411. Business Failure
412. Education
413. Professional Responsibilities
414. Forecasts
415. Decision Aids
416. Organization & Environment
417. Litigation
418. Governance

H. Accounting Area
1. Tax
2. Financial
3. Managerial
4. Audit
5. Information Systems
6. Mixed

I. Geography
1. Non-USA

2. USA
3. Both

J. Objective
1. Profit
2. Not for Profit
3. Regulated
4. All

K. Applicability
1. Immediate
2. Medium Term
3. Long Term

L. Foundation Discipline
1. Psychology
2. Sociology, Political Science, Philosophy
3. Economics & Finance
4. Engineering, Communications & Computer Sciences
5. Mathematics, Decision Sciences, Game Theory
6. Statistics
7. Law
8. Other Mixed
9. Accounting
10. Management

Appendix II

II.A. Description of taxonomy classes: accounting area, treatment and mode of reasoning

Accounting area	Identifies the major accounting field the paper covers. The major fields are tax, financial, managerial, audit, and information systems.
Treatment	Identifies the major factor or other accounting phenomena associated with or causing the information taxon. The treatment taxon will be the main predictor variable in the regression model in an empirical study. Main subcategories are financial accounting methods, auditing, managerial and others.
Mode of reasoning	Identifies the technique used to formally arrive at the conclusions of the study, either by quantitative or qualitative analysis. The quantitative subcategory includes various items, e.g. descriptive statistics, regression, ANOVA, factor analysis, non-parametric, correlations, and analytical.

Appendix III

III.A. Data mining algorithms

Classification 1: Bayes classifier
Algorithms
Bayesnet
DMNB Tex
Naïve Bayes
Naïve Bayes multinomial
Naïve Bayes multinomial updateable
Naïve Bayes updateable
Complement Naïve Bayes

Appendix III (*continued*)

Classification 2: decision trees classifier
Algorithms
J48
J48 graft
LAD tree
Random forest
Random tree
REP tree
Simple cart

Classification 3: rule based classifier
Algorithms
ZeroR
Ridor
PART
OneR
JRip
Decision table

Classification 4: other miscellaneous algorithms
Algorithms
Classification via
Regression
Multiclass classifier
Simple logistic
SMO
Attribute selected
Classifier
Bagging
Classification via
Clustering
CV parameter selection
Dagging
Decorate
END
Ensemble selection
Filtered classifier
Grading
Logit boost
Ensemble selection
Filtered classifier
Grading
Logit boost
Multi boost AB
Ensemble selection
Filtered classifier
Grading
Logit boost
Multi boost AB
Multi scheme
ND

Appendix IV

IV.A. Preliminary analysis I: detailed classification results.

Table 12
Results of Phase I Keywords Study – Treatment Taxon – with Five-fold Cross Validation.

12.1. Classification 1: Bayes classifier

Algorithm	Correctly classified
Bayes net	59.2%
Complement Naive Bayes	29.6%
DMNB Tex	59.2%
Naive Bayes	58.16%
Naive Bayes multinomial	56.12%
Naive Bayes multinomial updateable	56.12%
Naive Bayes updateable	58.16

12.2 Classification 2: decision trees classifier

Algorithm	Correctly classified
J48	56.12%
J48graft	56.12%
LADTree	57.14%
Random forest	57.14%
Random tree	55.1%
REP tree	59.18%
Simple cart	60.25%
LMT	56.12%
NB tree	59.18%

12.3 Classification 3: rule based classifier

Algorithm	Correctly classified
Zero R	59.18%
Ridor	59.18%
PART	55.1
One R	56.12%
JRip	57.14%
NNge	46.94%
Decision stump	57.14%
FT	54%

12.4 Classification 4: miscellaneous

Algorithm	Correctly classified
Logistic	58.16%
Simple logistic	56.12%
SMO	57.14
K Star	57.14
LWL	57.14
Attribute selected classifier	57.14
Bagging	59.18%
Classification via clustering	59.18%
Classification via regression	59.18%
CV parameter selection	59.18%
Dagging	59.18%
Decorate	56.12%
END	59.18%
Ensemble selection	59.18%
Filtered classifier	59.18%
Grading	59.18
Logit boost	55.1%
Multi boost AB	57.14%

Table 12 (*continued*)

12.4 Classification 4: miscellaneous

Algorithm	Correctly classified
Multiclass classifier	57.14%
Multi scheme	59.18%
Nested dichotomies class balanced ND	57.14%
Data near balanced ND	59.18%
ND	59.2%
Ordinal class classifier	59.18%
Raced incremental logit boost	59.18%
Random committee	58.16%
Random subspace	59.18%
Rotation forest	57.14%
Stacking	59.18%
Stacking C	59.18%
Vote	59.18%

Table 13
Results of Phase II – Treatment Taxon – Full Abstract Study.

13.1 Classification 1: Bayes classifier

Algorithm	Correctly classified
Bayes Net	69.5%
Complement Naive Bayes	74.01%
DMNB Text	68.36%
Naive Bayes	54.23%
Naive Bayes multinomial	72.88%
Naive Bayes updateable	54.24
SMO	58.75
Decision table	70.05

13.2 Classification 2: decision trees classifier

Algorithm	Correctly classified
J48	66.67%
J48 graft	71.75%
LAD tree	65%
Random forest	69.5%
Random tree	51.41%
REP tree	65.54%
Simple cart	70.05%
BF tree	70.06%
FT	63.84%

13.3 Classification 3: rule based classifier

Algorithm	Correctly classified
Zero R	48.6%
Ridor	63.28%
PART	62.71%
One R	65.54%
JRip	63.28%
Decision table	70.05%
Conjunctive rule	65.54%
NNge	58.76%
Decision stump	65.54%

13.4 Classification 4: miscellaneous

Algorithm	Correctly classified
SimpleLogistic	66.67%
SMO	58.76%

Table 14
Results of Phase II – Treatment Taxon – full abstract study with percentage split (66%).

14.1. Classification 1: Bayes classifier

Algorithm	Correctly classified
Bayes net	80%
Complement Naive Bayes	81.67%
DMNB text	68.33%
Naive Bayes	63.33%
Naive Bayes multinomial	80%
Naive Bayes updateable	63.34%
Naive Bayes multinomial updateable	80%

14.2 Classification 2: decision trees classifier

Algorithm	Correctly classified
J48	65%
J48 graft	73.33%
LAD tree	50%
Random forest	70%
Random tree	38%
REP tree	80%
Simple cart	75%
BF tree	64%
FT	65%
Decision stump	76.7%
LMT	70%

14.3 Classification 3: rule based classifier

Algorithm	Correctly classified
Zero R	53.33%
Ridor	61.67%
PART	68.33%
One R	77%
JRip	52%
Decision table	80%
NNge	60%

14.4 Classification 4: miscellaneous

Algorithm	Correctly classified
Classification via regression	78.33%
Multiclass classifier	43.33%
Simple logistic	70%
SMO	56.67%
Attribute selected classifier	78.33%
Bagging	78.33%
Classification via clustering	51.67%
CV parameter selection	53.33%
Dagging	68.33%
Decorate	78.33%
END	81.67%
Ensemble selection	78.33%
Filtered classifier	80%
Grading	54%
Logit boost	73.4%
Multi boost AB	76.67%
Multi scheme	53.33%

Table 15
Results of Phase I – accounting area – keywords study.

Algorithm	Correctly classified
15.1 Classification 1: Bayes classifier	
Bayes net	60.33%
Naïve Bayes	69.116
Naïve Bayes multinomial	69.116
Naïve Bayes multinomial updateable	69.116%
Complement Naïve Bayes	45.35%
15.2 Classification 2: decision trees classifier	
J48	68%
J48 graft	68%
Random forest	65.51%
Random tree	65.51%
Simple cart	64.43%
15.3 Classification 3: rule based classifier	
ZeroR	60%
PART	52%
JRip	61.32%
Decision table	60.2762%
Conjunctive rule	60.8287%
Ridor	54%

Table 16
Results of Phase II – Accounting Area – full abstract study with percentage split (66%).

Algorithm	Correctly classified
16.1 Classification 1: Bayes classifier	
Bayes net	83.2%
Complement Naive Bayes	85.31%
Naive Bayes	80.42%
Naive Bayes multinomial	71.36
16.2 Classification 2: decision trees classifier	
J48	74%
J48 graft	74%
Random forest	73%
Random tree	57%
16.3 Classification 3: rule based classifier	
ZeroR	73%
Ridor	71.36%
PART	71.36%
One R	71.36%
JRip	74.33%

Table 17
Results of Phase I — mode of reasoning keywords study.

17.1 Classification 1: Bayes classifier

Algorithm	Correctly classified
Bayes net	50.32%
Naïve Bayes	53.32%
Naïve Bayes multinomial	51.02%
Naïve Bayes multinomial updateable	54.62%
Complement Naïve Bayes	27%
DMNB text	53.32%

17.2 Classification 2: decision trees classifier

Algorithm	Correctly classified
J48	56.11%
J48graft	56.11%
Random forest	53.89%
Random tree	48.89%
Simple CART	54.44%
BFTree	55.55%
Decision stump	55.55%
REP tree	53.33%

17.3 Classification 3: rule based classifier

Algorithm	Correctly classified
Zero R	53.33%
PART	54.44%
JRip	54.16%
Decision table	54.16%
Conjunctive rule	56.42%
Ridor	52.22%
One R	56.62%

Table 18
Results of Phase II — mode of reasoning full abstract study.

18.1 Classification 1: Bayes classifier

Algorithm	Correctly classified
Bayes net	40%
Complement Naive Bayes	48.12%
Naive Bayes	46.69%
Naive Bayes multinomial	47.05%
DMNB text	48.13%
Naïve Bayes multinomial updateable	47.05%
Naïve Bayes updateable	46.69%

18.2 Classification 2: decision trees classifier

Algorithm	Correctly classified
J48	35.45%
J48 graft	36.61%
Random forest	42.35%
Random tree	31.58%
Simple CART	39.85%

18.3 Classification 3: rule based classifier

Algorithm	Correctly classified
Zero R	40.21%
Ridor	33.02%

Table 18 (*continued*)

18.3 Classification 3: rule based classifier	
Algorithm	Correctly classified
PART	34.46%
JRip	43.45%
Conjunctive rule	40.93%

References

Ashton RH. Human information processing in accounting. Studies in Accounting Research. Sarasota, Florida: American Accounting Association; 198217.

Badua FA. Pondering Paradigms: Tracing the Development of Accounting Thought with Taxonomic and Citation Analysis. Ph.D. DissertationRutgers University; 2005.

Badua F, Previts GJ, Vasarhelyi M. Tracing the development of accounting thought by analyzing content, communication, and quality in accounting research over time. Account Hist J 2011;38:31–56.

Ball R. Index of empirical research in accounting. J Account Res 1971;9:1–31.

Barki H, Rivard S, Talbot J. A keyword classification scheme for is research literature: an update. MIS Q 1993;17:209–26.

Birnberg JG, Shields JF. Three decades of behavioral accounting research: a search for order. Behav Res Account 1989;1:23–74.

Brown LD, Gardner JC, Vasarhelyi MA. An analysis of the research contributions of accounting, organizations and society, 1976–1984. Acc Organ Soc 1987;2:193–204.

Brown LD, Gardner JC, Vasarhelyi MA. Attributes of articles impacting contemporary accounting literature. Contemp Account Lit 1989;5:793–815.

Brown LD, Gardner JC, Vasarhelyi MA. Accounting research directory: the database of accounting literature. New York: Marcus Wiener Publishing; 1994.

Chatfield M. The accounting review's first fifty years. Account Rev 1975;1:1–6.

Chen H, Lynch KJ. Automatic construction of networks of concepts characterizing document databases. IEEE Trans Syst Man Cybern 1992;22:885–902.

Chen H, Schatz B, Martinez J, Ng TD. Generating a domain-specific thesaurus automatically: an experiment on FlyBase. Working Paper MAI-WPS 94–02: Center for Management of Information College of Business and Public Administration, University of Arizona; 1994.

Chen H, Yim T, Fye D. Automatic thesaurus generation for an electronic community system. J Am Soc Inf Sci 1995;46:175–93.

Crawford RG. Automatic thesaurus construction based on term centroids. Can J Inf Libr Sci 1979;4:124–36.

Crouch CJ. An approach to the automatic construction of global thesauri. Inf Process Manag 1990;26:629–40.

Crouch CJ, Yang B. Experiments in automatic statistical thesaurus construction. Paper presented at the Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Copenhagen Denmark; 1992.

Dyckman TR, Zeff SA. Two decades of the journal of accounting research. J Account Res 1984;22:225–97.

Felix WL, Kinney WR. Research in the auditor's opinion formulation process: state of the art. Account Rev 1982;57:245–71.

Fisher IE, Garnsey MR, Goel S, Tam K. The role of text analytics and information retrieval. J Emerg Technol Account 2010;7:1–24.

Fleming RJ, Graci SP, Thompson JE. The dawning of the age of quantitative/empirical methods in accounting research: evidence from the leading authors of the accounting review, 1966–1985. Account Hist J 2000;27:43–72.

Gangolly J, Wu Y-F. On the automatic classification of accounting concepts: preliminary results of the statistical analysis of term-document frequencies. New Rev Appl Exert Syst Emerg Technol 2000;6:81–8.

Garnsey MR. Automatic classification of financial accounting concepts. J Emerg Technol Account 2006;3:21–39.

Garnsey MR, Fisher IE. Appearance of new terms in accounting language: a preliminary examination of accounting pronouncements and financial statements. J Emerg Technol Account 2008;5:17–36.

Gonedes N, Dopuch N. Capital market equilibrium, information production, and selecting accounting techniques: theoretical framework and review of empirical work. J Account Res Suppl 1974:48–170.

Hakansson N. Empirical research in accounting, 1960–70: an appraisal. In: Dopuch N, Revsine L, editors. Accounting Research: 1960–1970: A Critical Evaluation. Urbana: Center for International Education and Research in Accounting, University of Illinois; 1973. p. 137–73.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The Weka data mining software: an update. SIGKDD Explor 2009;11.

Heck JL, Jensen RE. An analysis of the evolution of research contributions by *the accounting review*, 1926–2005. Account Hist J 2007;34:109–41.

Hofstedt TR. A state-of-the art analysis of behavioral accounting research. J Contemp Bus 1975(Autumn):27–49.

Hofstedt TR. Behavioral accounting research: pathologies, paradigms and prescriptions. Acc Organ Soc 1976;1:43–58.

Hulth A. Improved automatic keyword extraction given more linguistic knowledge. Proceeding EMNLP '03 proceedings of the 2003 conference on empirical methods in natural language processing; 2003.

Ijiri Y, Kinard JC, Putney FB. An integrated evaluation system for budget forecasting and operating performance with a classified budgeting bibliography. J Account Res 1968;6:1–28.

Just A, Meyer M, Schaffer U, Vasarhelyi MA, Chiu V. Promoting diversity in accounting research: an empirical study of the development of the identity and intellectual structure of AOS from 1990 to 2007. Working Paper; 20121–45.

Krippendorff K. Content analysis. Thousand Oaks, CA: Sage; 2004.

Libby R, Lewis BL. Human information processing research in accounting: the state of the art. Acc Organ Soc 1977;2:245–68.

Libby R, Lewis BL. Human information processing research in accounting: the state of the art in 1982. Acc Organ Soc 1982;7:231–85.
Liu B, Hsu W, Ma Y. Integrating classification and association rule mining. KDD'98; 1998.
March ST, Smith GF. Design and natural science research on information technology. Decis Support Syst 1995;15:251–66.
Meyer M, Rigsby JT. A descriptive analysis of the content and contributors of behavioral research in accounting 1989–1998. Behav Res Account 2001;13:253–78.
Nigam K, Miccallum AK, Thrun S, Mitchell T. Text classification from labeled and unlabeled documents using EM. Mach Learn 2000;39:103–34.
Nobata C, Collier N, Tsu J. Automatic term identification and classification in biology texts. Proceedings of the Natu'ral Lang,Imgc Pacific Rim Symposium; 1999.
Previts GJ, Brown RE. The development of government accounting: A content analysis of the Journal of Accountancy 1905–1989. Account His J 1993;20:119–38.
Salton G, Allan J, Singhal A. Automatic text decomposition and structuring. Inf Process Manage 1996;32:127–38.
Tan P, Steinbach M, Kumar V. Introduction to data mining. Addison Wesley; 2005.
Thompson CA, Califf ME, Mooney RJ. Active learning for natural language parsing and information extraction. Proceedings of the International Conference on Machine Learning (ICML); 1999406–14.
Trueswell JC, Tanenhaus MK, Garnsey SM. Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. J Mem Lang 1994;33:285–318.
Tucker AB. A perspective on machine translation: theory and practice. Commun ACM 1984;27:322–9.
Vasarhelyi MA, Bao DH, Berk J. Trends in the evolution of scholarly accounting thought: a quantitative examination. Account Hist J 1988;15:45–64.