Improving Identification Of Search Terms In The FARS Database

Margaret R. Garnsey, (Email: garnsey@siena.edu), Siena College
Andrea W. Hotaling, (Email: hotaling@siena.edu), Siena College

ABSTRACT

The diversity of language and the evolving nature of accounting terminology make it difficult for users to generate adequate query terms when doing keyword searching in the authoritative accounting literature. This project demonstrates one method of improving identification of retrieval query terms by presenting searchers with lists of terms specific to the Financial Accounting Research System (FARS) database. The results show individuals improve both the number and relevance of possible search terms when provided with lists of potentially appropriate terms.

INTRODUCTION

wo developments in recent years have exacerbated the already difficult information retrieval process when searching authoritative accounting literature. The first trend is simply the proliferation of accounting pronouncements and terminology and the ever-increasing complexity of the issues addressed by the Financial Accounting Standards Board (FASB). A Special Task Force on Standards Overload of the AICPA defined standards overload, in part, as "the rapid change in and expansion of performance and reporting standards" (Burke, 1997).

The second trend is the movement towards online access and keyword searching of the authoritative sources. Based on their research with full-text document-retrieval systems, Blair & Maron (1985) state, "A full-text retrieval system…places the user in the position of having to find an impossibly difficult combination of search terms." Furnas et al (Furnas, Landauer, Gomez, & Dumais, 1987) documented this vocabulary problem in full text searching with their study that showed that users in five different domains favored the same spontaneously generated terms with a probability of less than 0.20. Similarly, Foltz highlighted the deficiency of keyword matching "People seldom know which words will describe a document and there is a great variability in the choice of words between people" (Foltz, 1991). Beyond the issues caused by the pure vocabulary problem, Blair & Maron (1985) illustrated how full text retrieval systems create output overload much quicker than manually indexed systems and how the addition of search terms to reduce the output often eliminates relevant documents.

Authoritative Accounting Literature And Keyword Searching

In an evolving technical field such as accounting research skills are a necessity to stay abreast of the current authoritative base. Accounting students, in particular, have an increased need to master research skills. The recently promulgated AICPA Core Competency Framework identifies research skills as one of the necessary functional competencies that an entry level accounting professional should possess. Additionally, the revisions to the newly adopted computer based CPA exam have added simulation type problems to assess those research skills.

FARS (Financial Accounting Research System) is an online full text document collection for authoritative accounting literature used in practice and as the source for the accounting simulations on the computerized CPA exam. All of the difficulties of full text searching highlighted above apply to the keyword searching in FARS. Enhancements to the search functions within this authoritative accounting literature have the potential to assist all users in improving both the retrieval and relevance of their search results but would be particularly valuable to students and other inexperienced researchers.

Folio Views is the commercial software package whose interface provides the search functions used to access the FARS database of authoritative literature. Folio Views basic search function offers the usual Boolean search techniques and like most conventional retrieval techniques, requires the matching of terms in the query to the terms in a document. This requirement often leads to unsatisfactory results because of the required exact match between search terms and the terms used in documents.

Folio Views does offer an advanced search feature where the initial query terms can be automatically augmented by a list of synonyms. However, although the Folio Views synonym list provides synonyms for individual terms it does not provide synonyms for entire phrases which are often the keywords needed for a search. Another weakness of the Folio Views synonym list is that it is a generic list without an accounting context so in fact it may create difficulties of output overload that more than offset the benefits of additional synonyms. For example, the term security readily brings to mind a certain concept for accountants. The automatic synonym list brings up 28 terms (see Table 1) including the terms: safety, assurance, and conviction. The list does not include the terms stock or shares which, in many cases, accountants would consider to be more relevant than the terms on the list. The addition of the synonyms expands the hit list by ~40%. Given that most of the synonyms are not relevant in an accounting sense, the results would now include a higher proportion of irrelevant documents.

Table 1: Folio Views Synonyms For The Term "Security"				
securities	safeguard	conviction	deposit	warranty
safety	preservation	certainty	forfeit	token
protection	confidence	soundness	pledge	bond
shield	assurance	faithfulness	promise	escrow
defense	collateral	surety	contract	warrant
shelter	freedom from harm			

Our research illustrates how a list of potential search terms or phrases automatically generated from the FARS corpus (the authoritative literature included in FARS), can improve the ability of searchers to identify relevant terms and phrases and accordingly has the potential to improve the effectiveness of their research. We empirically tested the improvement in both the number and relevance of possible search terms identified by a group of accounting educators when they were provided with lists of terms from a domain specific thesaurus created by Garnsey (2001).

LITERATURE REVIEW

The Vector Space Model In Retrieval

One model of information retrieval is the vector space model which was originally developed by Salton (Salton, G., Wong, A., & Yang, C. S., 1975). The model represents both queries and documents as term-vectors in a *t*-dimensional space, where *t* represents the number of terms in the vocabulary. The similarity between any two vectors can be computed as the inner product between them.

In an expansion of the vector space model, Salton, Yang, & Yu (1975) define the discrimination value of a term as a measure of the changes in separation between documents in a t-dimensional space which occurs when a given term is assigned to a collection of documents. A good discriminator is one which will render documents less similar to each other, reducing the likelihood of irrelevant documents being returned for a given query. The authors have shown that good discriminators and thus better retrieval terms are terms that do not occur as often (have a lower frequency) in a document collection. Salton, Yang et al (1975) demonstrated that use of the discrimination value, in term weighting, improves both recall (the number of relevant documents retrieved) and precision (the proportion of relevant documents to all documents retrieved). In analyzing their results, they found that "the terms with high document frequencies…are the worst discriminators." A system that could generate a suggested list of lower frequency terms would be an enhancement to the retrieval process.

Improving Identification Of Possible Retrieval Terms

The retrieval of information involves three stages (Foltz, 1991): generating retrieval cues (search or query terms), using them to retrieve information, and evaluating whether the retrieved information is what is desired. The first step of the retrieval process requires the user to generate terms pertinent to his/her question. The experiences and personal history of people lead them to use different terms to describe similar concepts. This circumstance can result in it being difficult for a user to predict what terms authors of relevant documents have used, leading to a failure to obtain all relevant documents.

Spink & Saracevic (1997) looked at the source and effectiveness of search terms "under real-life (as opposed to laboratory) circumstances." Their subjects included forty self-selected faculty and doctoral students with real information problems who provided one question each for on-line searching on DIALOG. Spink & Saracevic's research found that only half of the relevant items retrieved were found using only terms supplied by the user.

A retrieval thesaurus or some type of term list is often used in library/information science so the user and indexer have access to same terms. Most thesauri are for a specific discipline reflecting the terms used in the literature on a subject. Muddamalle (1998) has noted that in fields where language is changing a thesaurus needs to be regularly updated but it is difficult to answer the question of how often this should be done. His research found that only 45% of request terms matched with terms in the *Microthesaurus of Soil Mechanics* (published in 1974).

Price Waterhouse & Co., one of the big four accounting firms, published an accounting thesaurus, for internal use, in 1974 (Price Waterhouse & Co., 1974), however, due to the evolution of accounting language many terms that are common today are not included in it (ex: functional currency, derivatives). Even this limited resource is not widely available forcing people to generate their own query terms when doing research in FARS.

Automatic Generation Of Related Terms

Library and Information Science researchers have looked at using vocabulary-based search aids in an on-line environment where documents have not been manually indexed. The majority of this research has been done on experimental collections. In these test collections, the relevant documents for assessment queries are known allowing the traditional measures of retrieval performance (recall: the number of relevant documents returned and precision: the proportion of returned documents that are relevant) to be used. The research described below investigates the automatic construction of domain-specific thesauri using the terms present in real-life document collections where the relevant documents to test queries is not known.

In the scientific realm, Chen & Lynch (1992) derived an algorithm, based on the probability of term co-occurrence in document sets, to identify related terms. Chen, Yim, & Fye (1995) used this approach to automatically generate a thesaurus for the Worm Community System (WCS). The WCS contains a comprehensive library of specialized community data and literature in use by molecular biologists who study the nematode worm, C.elegans. The authors found a statistical difference (at the 10% confidence level) in the mean number of terms generated by subjects in a recall test and the terms generated using their algorithmic approach. Similar results were obtained using the same approach to create a domain-specific thesaurus for Drosophila (fruit fly) information (Chen, Schatz, Martinez, & Ng, 1994) and for a large, operational textual database of computing knowledge (Chen & Lynch, 1992).

Garnsey (2001) developed a technique that combined Latent Semantic Indexing (LSI) and agglomerative clustering to derive clusters of related accounting concepts using the authoritative literature in FARS. LSI is a method of two mode factor analysis that infers relationships between terms that are used in the same context, even if the terms have never been used together in the same document. Agglomerative clustering begins with each term as a separate cluster and successively merges the two most similar clusters until all terms have been merged. Garnsey used these techniques to statistically derive sets of clusters of accounting terms contained in the documents of the FARS database. The clusters created in this manner were able to identify significantly more terms relevant to a given test term with a higher precision (relevant terms in a cluster compared to all terms in a cluster), ~80%, than selecting terms at random.

Limitations Of Authoritative Accounting Literature

Several scholars have discussed the shortcomings in the authoritative accounting literature. Gangolly, Hedley, & Wong (1991) stated "The accounting standards …exhibit ambiguity, semantic richness, and extremely complicated rules for inference…"   They proposed using an object-oriented language, such as PROLOG (which encourages declarative programming) to represent the rules in accounting in order to develop a knowledge base. This would require an in-depth semantic analysis of the accounting standards which has not been undertaken at the present time.

More recently, Fisher (2003) states: " Financial accounting standards, as currently drafted, lack a consistent and sound structural arrangement and are incompatible with digital storage and retrieval methods." She makes suggestions for restructuring Financial Accounting Standards to facilitate the application of digital techniques as well as to enhance the understandability of standards. Her proposal includes the use of mark-up languages to formally describe and standardize the structure of accounting standards through the use of DTD's (Data Type Definitions) or similar methods.   Neither of these proposals has advanced beyond the theoretical stage. As such, we see the need to make improvements in the existing search process.

HYPOTHESIS DEVELOPMENT

Helping Users Identify Potential Retrieval Terms

Recall is the ability to remember information without being given any cues. Recognition is the awareness that something perceived has been perceived before. Previous experiments, in psychology, demonstrate that recognition is generally better than recall. Higher recognition may occur because a recognition test typically provides more sources for activating memory (Anderson, 1990), 172). Gibson, Bishop, Schiff, & Smith (1964) found that recognition was higher than recall for trigrams. This was true no matter whether the trigrams were meaningful (common initials), pronounceable (obeying rules of English grapheme-phoneme correspondence) but not meaningful, or neither pronounceable nor meaningful.   Nelson (1978) had subjects learn number-word pairs. He found that recognition of the number-word pair was substantially better than recall of the word when given the number. The above research provides evidence that recognition is better than recall in various experimental settings. We expand this research to a real-life accounting setting.

As described in the literature review, Chen's work dealt with scientific databases. Our work uses the technique developed by Chen to examine users' recall in an accounting database. When trying to expand a query users may "know" more related terms but be unable to recall them because the connections to them, in memory, are too weak. However, by providing an individual with a list of terms he/she may be able to identify additional related terms, which would aid in retrieval. Our research asked subjects, for given test terms, to generate or identify possible retrieval terms in two ways: first, by recalling terms from memory and second through recognition when given a list of terms taken directly from FARS and used in similar contexts.

H1:

Providing users with a group of terms from the vocabulary used in the FARS database will allow them to identify more relevant terms and generate better concept recall than they are able to from memory alone.

Automatically Grouped Terms Have Lower Frequency

In information retrieval, people tend to use terms that have a high frequency of occurrence (Peat & Willett, 1991). This is in agreement with research in human memory which has indicated that persons are more likely to recall words that occur frequently (Kintsch, 1977). Previously, Salton, Yang et al (1975) found that the worst terms for retrieval are those which occur in the highest number of documents. In the FARS database using the word "financial" produces 6,614 hits, most of which would not be useful in a particular search.

Good retrieval terms were found to be terms that have a lower frequency in a document collection Salton, Yang et al (1975). Individuals can often recognize these infrequent terms even when they are unable to produce the terms through memory recall. In a recognition experiment that involved common words (occurring at least 100 times per million sampled words) and rare words (occurring less than once per million), Shepard (1967) found that subjects had higher recognition for rare words than words which were common.

The above research shows that terms with a lower frequency are better candidates for search queries and individuals can often recognize these terms even when they cannot recall them. By providing users with a list of potential retrieval terms they can recognize relevant lower frequency terms, which could then be used to expand their queries.

H2:

Relevant terms, which were automatically generated from the FARS database, will have a lower frequency than terms generated by individuals from memory alone.

METHODOLOGY

Participants

The participants selected included accounting faculty from several institutions. As indicated in Table 2 most subjects had both substantial accounting and academic experience (27.2 and 20.6 average years respectively). In addition all but two participants were licensed CPAs, two were CMAs and three held Ph.Ds. This group of individuals represents a sample of individuals with exposure to accounting pronouncements and presumed familiarity with accounting terminology.

Table 2: Characteristics Of Subjects									
	s01	**s02**	**s03**	**s04**	**s05**	**s06**	**s07**	**Total**	**Average**
CPA	x	x	x	x			x	5	
CMA						x	x	2	
Ph.D.					x	x	x	3	
yrs in accounting	18	32	40	26	26	26	26		27.2
yrs teaching	4	29	34	20	16	26	15		20.6

Design

This research used clusters of terms derived from the FARS database (Garnsey, 2001) to provide lists of potential terms to the subjects in the recognition test. The clusters were created using the vocabulary present in official accounting pronouncements. Chen (Chen & Lynch, 1992) developed a method to evaluate automatically generated thesauri using actual document collections from the scientific realm. In this research, we followed his methodology in the accounting realm. Data was collected in two phases. Phase I involved a term association or recall test. In Phase II, performed approximately two weeks after phase I, a recognition test was performed. The time lag was necessary to allow decay in memory from the recall test.

Procedure

The thirty terms used in the recall and recognition tests were selected at random from the vocabulary used by Garnsey (2001) to generate term clusters. In phase I, the term association or recall test, subjects were asked to recall as many related concepts as possible for each of the thirty terms, without being given any additional clues or synonyms. See Figure 1 adopted from (Chen et al., 1994).

Figure 1: Form For Subject-Suggested Descriptors

Please write as many terms, phrases, and concepts as possible that are related to each given term. A term is considered related if it has a similar meaning to the given term, or if it is associated with the same subject as the given term, or the terms would be used together. Include any term that would be considered somewhat relevant. Please do not guess. If you do not know a term, leave it blank.	
Combination	

In Phase II, a recognition test was performed. The subjects were given the same thirty test terms together with an associated list for each test term and were asked to evaluate each word/phrase on the list according to the following scale: "Irrelevant," "Somewhat Relevant," "Very Relevant." Words/phrases that were too general or that the subjects were unfamiliar with were to be ranked as "Irrelevant." (see Figure 2, adopted from Chen et al., 1994). The list for each test term contained terms from the clusters generated by Garnsey (2001) as well as the suggested terms given by all subjects in the term association test.

Figure 2: Sample System-Suggested Descriptors

Please evaluate each term in the list of terms for its relevance or association to the term in bold type at the top of the column by checking the appropriate box. A term is considered relevant or associated if it has a similar meaning to the given term, or if it is related to the same subject as the given term, or it would be used in discussing the given term. If you are unsure of a term's relevance please mark it as irrelevant. *A term that is too general is considered irrelevant.*			
Combination			
Irrelevant	Somewhat Relevant	Very Relevant	
			subsidiary
			impairment
			maturity
			investor
			premium
			portfolio
			segment
			etc….

Following the procedure developed by Chen (Chen & Lynch, 1992), we compared the concept recall of the subjects in the term association test with their concept recall in the recognition test. Concept recall is a modification of one of the traditional information retrieval measures of performance. It compares the relevant concepts identified by the subjects to the total relevant concepts.

RESULTS

Summary

 A summary of the results is given in Table 3. In the term association test, the subjects were able to generate an average of 5.1 words/phrases per test term as possible relevant terms. In the recognition test, the subjects were able to identify an average of 45.8 terms as "very relevant" or "somewhat relevant." The provision of the list of suggested terms from other participants and the automatic clustering increased nine fold the possible relevant terms.

Table 3: Summary of Responses								
	s01	s02	s03	s04	s05	s06	s07	All s's
Term Association Test								
total responses	67	83	129	147	117	211	194	948
avg. responses per term	2.6	4.0	4.4	5.3	4.9	7.8	6.9	5.1
Recognition Test								
Very or Somewhat Relevant Terms	1,084	1,196	927	1,394	1,150	1,336	2,043	1034
avg. responses per term	40.1	40.2	35.7	46.5	40.0	47.7	70.4	45.8

The average overlap of words/phrases given by subjects in the term association test averaged approximately 12%. This is similar to the results demonstrated by prior research in human-computer interaction where the probability of two persons using the same terms for an object averaged less than 20% (Furnas et al., 1987). Word/phrase overlap between the terms recalled by the subjects vs. the terms suggested by the clusters was also low. Total overlap for all terms given by the subjects and the terms identified as relevant from the clusters was less than 4%. This is similar to the results obtained by Chen & Lynch (1992) for a computing database, where they found that term overlap between their automatically identified terms and the terms recalled by subjects ranged from 4.9% to 8.6%. The clusters were able to suggest a significant number of relevant words/phrases beyond terms identified collaboratively by subjects.

Tests Of Hypotheses

The results from the Term Association Test and Recognition Test were compared. To determine if subjects were able to identify additional relevant words/phrases when presented with a list the following null hypothesis (H1) was tested collectively and for each subject individually:

Providing users with a group of terms from the vocabulary used in the FARS database will not allow subjects to identify more terms or generate better concept recall than they are able to do from memory alone.

Initially a paired t-test was used to test the hypothesis. The null hypothesis was rejected, $p < 0.01$. Because the t-test assumes a normal distribution, a non-parametric test, Wilcoxon signed-rank test, was also used. This confirmed the results from the t-test, the null hypothesis was rejected, $p < 0.01$.

The data demonstrates that the subjects were able to recognize a significantly higher number of terms in the recognition test than they were able to recall in the term association test. The boxplots in Figure 3 give an overall view of the data. The median is shown by the solid white line in the box with the ends of the box marking quartiles. The 'whiskers' show the extremes of the data and outliers are shown by themselves. In the term association test, the subjects recalled an average of 5.1 words/phrases per test term. They recognized an average of 45 words/phrases in the recognition test as very or somewhat relevant, an average increase of 900%. The fact that recognition was higher than recall agrees with previous research on automatic thesaurus construction (Chen & Lynch, 1992) and research on memory (Gibson, Bishop, Schiff, & Smith, 1964; Nelson, 1978).

The frequency of a term is the number of times it appears in the document collection. The frequency of terms subjects gave, that were included in the final vocabulary, was compared to the frequency of the relevant terms from each set of clusters. The following null hypothesis (H2) was tested:

Relevant terms which were automatically generated from the FARS database will have the same average frequency as terms generated by individuals from memory alone.
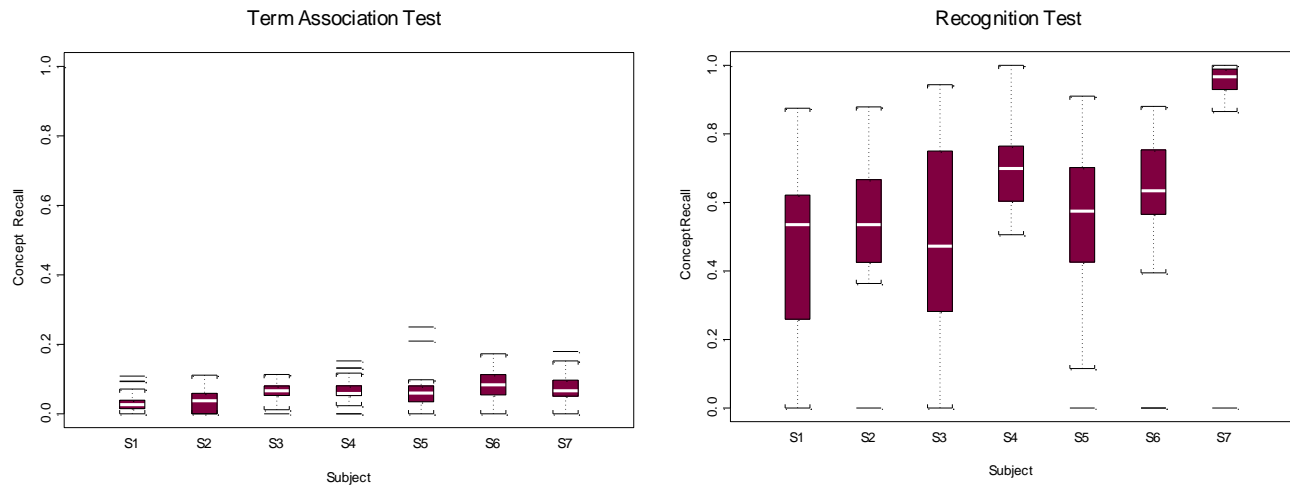
Term Association Test   Recognition Test

Figure 3: Boxplot Comparison of Subject's Concept Recall

Using the Wilcoxon rank sum test, the null hypothesis was rejected (p< 0.001). The terms identified by the subjects in the term association test had a significantly higher average frequency in the accounting pronouncements, 207 times, than the automatically generated cluster terms identified in the recognition test, 95 times.

DISCUSSION AND CONCLUSIONS

The total number of unique relevant words/phrases suggested by the subjects in the term association test was 603. Of these terms, 216 were included in the vocabulary used in generating the clusters, 27 were eliminated from the original vocabulary due to their frequency (either greater than 1500 appearances or appearance in less than 3 documents), 320 appear in the FARS database but were not included in the vocabulary, and 66 did not appear in the FARS database.

Of the relevant subject terms included in FARS but not in the original vocabulary, 102 would have been eliminated from the vocabulary due to their frequency in the accounting pronouncements. Fifty-four of the remaining terms were single words. Many of these had been eliminated in forming the vocabulary because they were not specifically related to accounting (Ex: month, annual, officers). The final 218 terms were phrases. While a small number of these phrases were not related uniquely to accounting (Ex: short term, for the period, period of time) the majority were directly associated with accounting. The average frequency of participant suggested phrases, 113, is significantly lower than the average frequency of the other relevant participant terms and would probably be useful in retrieval.

When searching in an online environment the user must be able to generate query terms that will match terms used in the relevant documents. This is a difficult task due to the diversity in vocabulary used. In accounting, this task is especially difficult due to the variety of users and the evolving nature of accounting language. For thirteen of the 30 terms selected at random for testing, there was at least one subject who could not generate any related terms in the term association (recall) test (see Table 4).

Table 4: Count Of Terms With No Response By Term			
Terms	**Term Frequency in Pronouncements**	**Subjects with no response in Recall Test**	**Subjects with no response in Recognition Test**
Qualifying spe	15	6	5
Feedback value	8	4	1
Look back options	9	4	3
Recognition of liabilities in connection with purchase business combination	1	3	
Recorded investment in the receivable	95	2	1
Hedge of net investment	28	2	
Right of return	115	1	
Gain contingencies	17	1	
Catch up adjustment	23		1
Non compensatory plans	26	1	1
Restructure	8	1	
Voting securities	8	1	
Income taxes on undistributed earnings	51	1	

However, the number of terms with no response declined significantly (using the Wilcoxon signed-rank test, $p = 0.002$) in the recognition test. In retrieval, if an initial query did not give satisfactory results a user who could not generate more query terms would remain unsatisfied. By presenting the user with other possibly relevant terms, he/she may be able to identify additional appropriate terms, which will increase his/her chances of getting satisfactory results.

Table 5: Count of Terms with No Response by Subject							
Subject	S01	S02	S03	S04	S05	S06	S07
Term Association Test							
Terms with no response	4	9	1	2	6	3	2
Recognition Test							
Terms with no response	3	1	4	0	1	2	1

In one case, S03, an individual who gave responses in the term association test did not give a response in the recognition test (see Table 5). This is despite the fact that, with one exception, responses the subject had given in the term association test were marked as "relevant" by the other subjects. It is thought that the subjects were less likely to guess in the recognition test, where they were instructed to mark terms they were unsure about as irrelevant, than in the term association test, where they were instructed to give as many terms as possible related to the given term.

This research shows that clusters of accounting concepts derived from the terms actually used in the authoritative literature in the FARS database can provide users with a set of candidate terms. If such an addition to the search process were incorporated into the professional literature it would provide users with a way of identifying additional retrieval cues, which could improve access to that literature. As such, this research is seen as an initial step in implementing the Special Task Force's recommendation for implementation of a system to facilitate access to and to improve the understanding of the professional literature (O'Dell, 1998).

Future Research Questions

Research has found that people had a higher success rate in finding targeted information objects as the number of index terms used for each object increased (Gomez, Lochbaum, & Landauer, 1990).   Tests comparing subjects with and without access to cluster terms when doing research in FARS needs to be conducted to determine if access to clusters increases efficiency of searching and completeness of results. A preliminary experiment has shown

that the usefulness of providing a list to students using FARS may depend on the research question being asked (Garnsey & Stokes, 2004).

The only phrases included in the vocabulary used by Garnsey (2001) were those found in the indexes of FARS and several accounting texts. The subjects used approximately 218 phrases which were not included in the vocabulary used in generating the clusters. Many of these phrases would be considered appropriate accounting language ("par value method," "like kind exchanges," "premium on notes payable"). Different statistical methods have been used to determine which words that occur together are collocations, meaningful phrases (Manning & Schutze, 1999). The use of automatic methods to identify phrases has been used in other research on developing concept spaces (Chen & Lynch, 1992; Chen, Ng, Martinex, & Schatz, 1997; Chen et al., 1995). By including appropriate collocations (terms which occur next to each other which are meaningful phrases), determined directly from the accounting pronouncements, it may be possible to significantly expand the number of meaningful phrases included in the vocabulary.

Finally, this research indicates that the method developed by Garnsey (2001) could be helpful in developing an up-to-date accounting thesaurus. The 1974 Price Waterhouse internal accounting thesaurus (Price Waterhouse & Co., 1974)is outdated given the proliferation of new pronouncments and terminology that did not exist in 1974 The clustering technique tested here may provide a starting point in developing an updated thesaurus function that could be applied to the accounting literature.

REFERENCES

1.   Anderson, J. R. (1990). *Cognitive Psychology and Its Implications* (3rd ed.). New York: W.H. Freeman and Co.
2.   Blair, D. C., & Maron, M. E. (1985). An Evaluation of Retrieval Effectiveness for a Full-Test Document-Retrieval System. *Communications of the ACM, 28*(3), 289-299.
3.   Burke, J. F. (1997). Report on Standards Overload. *CPA Journal, 66*(3), 11.
4.   Chen, H., & Lynch, K. J. (1992). Automatic Construction of Networks of Concepts Characterizing Document Databases. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(5), 885-902.
5.   Chen, H., Martinez, J., Kirchhoff, A., Ng, T. D., & Schatz, B. R. (1998). Alleviating Search Uncertainty through Concept Associations: Automatic Indexing, Co-Occurrence Analysis, and Parallel Computing. *Journal of the American Society for Information Science, 49*(3), 206-216.
6.   Chen, H., Ng, T. D., Martinex, J., & Schatz, B. R. (1997). A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System. *Journal of the American Society for Information Science, 48*(1), 17-31.
7.   Chen, H., Schatz, B., Martinez, J., & Ng, T. D. (1994). *Generating a Domain-Specific Thesaurus Automatically: An Experiment on FlyBase* (No. Working Paper MAI-WPS 94-02): Center for Management of Information College of Business and Public Administration, University of Arizona,.
8.   Chen, H., Yim, T., & Fye, D. (1995). Automatic Thesaurus Generation for an Electronic Community System. *Journal of the American Society for Information Science, 46*(3), 175-193.
9.   Crouch, C. J. (1990). An Approach to the Automatic Construction of Global Thesauri. *Information Processing and Management, 26*(5), 629-640.
10.   Crouch, C. J., & Yang, B. (1992). *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.
11.   Fisher, I. E. (2003). On the Structure of Financial Accounting Standards to Support Digital Representation, Storage and Retrieval. *Journal of Emerging Technologies in Accounting, 1*(1).
12.   Foltz, P. W. (1991, August 7-10, 1991). *Modeling Human Memory Retrieval and Computer Information Retrieval: What can the two fields learn from each other?* Paper presented at the Thirteenth Annual Conference of the Cognitive Science Society, Chicago, Illinois.
13.   Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The Vocabulary Problem in Human-System Communication. *Communications of the ACM, 30*(11), 964-971.

14.	Gangolly, J. S., Hedley, T. P., & Wong, C.-T. (1991). Semantic Knowledge Bases for Financial Accounting Standards. *Expert Systems with Applications, 3*, 117-128.
15.	Garnsey, M. R. (2001). *Classification of Financial Accounting Concepts Through the Use of Latent Semantic Indexing and Clustering Techniques.* University at Albany, State University of New York, Albany.
16.	Garnsey, M. R., & Stokes, L. (2004, August 8-11, 2004). *Finding the Right Word: A Preliminary Experiment in Providing Potential Terms to Students for FARS Searches.* Paper presented at the American Accounting Association Annual Meeting, Orlando.
17.	Gibson, E. J., Bishop, C. H., Schiff, W., & Smith, J. (1964). Comparison of Meaningfulness and Pronunciability as Grouping Principles in the Perception and Rention of Verbal Material. *Journal of Experimental Psychology, 67*(2), 173-182.
18.	Gomez, L. M., Lochbaum, C. C., & Landauer, T. K. (1990). All the Right Words: Finding What You Want as a Function of Richness of Indexing Vocabulary. *Journal of the American Society for Information Science, 41*(8), 547-559.
19.	Kintsch, W. (1977). *Memory and Cognition*. New York: John Wiley & Sons.
20.	Muddamalle, M. R. (1998). Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study in Soil Mechanics. *Journal of the American Society for Information Science, 49*(10), 881-887.
21.	Nelson, T. O. (1978). Detecting Small Amounts of Information in Memory: Savings for Nonrecognized Items. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 453-468.
22.	O'Dell, J. H. (1998). Accounting & Auditing. *Pennsylvania CPA Journal, 68*(4), 6-7.
23.	Peat, H. J., & Willett, P. (1991). The Limitations of Term Co-occurrence Data for Query Expansion in Documental Retrieval Systems. *Journal of the American Society for Information Science, 42*(5), 378-383.
24.	Price Waterhouse & Co. (1974). *Thesaurus of Accounting and Auditing Terminology*. New York: Price Waterhouse & Co.
25.	Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. In K. Sparck Jones & P. Willett (Eds.), *Readings in Information Retrieval* (pp. 272-280). San Francisco: Morgan Kaufmann Publishers, Inc.
26.	Salton, G., Yang, C. S., & Yu, C. T. (1975). A Theory of Term Importance in Automatic Text Analysis. *Journal of the American Society for Information Science, 26*(1), 33-44.
27.	Shepard, R. N. (1967). Recognition Memory for Words, Sentences, and Pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156-163.
28.	Spink, A., & Saracevic, T. (1997). Interaction in Information Retrieval: Selection and Effectiveness of Search Terms. *Journal of the American Society for Information Science, 48*(8), 741-761.

NOTES