USING TWO CONCEPT MAPPING TOOLS IN COMBINATION TO ANALYZE AND ORGANIZE DATA IN LARGE PUBLICATIONS

Anita Samuel PhD[1], Simone Conçeição PhD[2], & Larry Martin PhD[2]
[1]Uniformed Services University of Health Sciences, USA
[2]University of Wisconsin-Milwaukee, USA
Email: anita.samuel.ctr@usuhs.edu, {simonec, lmartin}@uwm.edu

Abstract. Analyzing and organizing large digital data sets can be overwhelming and time consuming. Concept mapping tools such as CmapTools and Leximancer allow for making meaning and connections, and synthesizing information and research data in a unique way. This paper presents an exploratory comparative analysis case study that implemented two approaches using these two concept mapping tools in combination, with the purpose of analyzing and organizing a large publication. The strengths and limitations of each approach are explored; and, suggestions for replicating the use of the two concept mapping tools in combination in other contexts are presented.

Keywords*:* CmapTools, Leximancer, big data, data analysis, large publications, concept mapping

1 Introduction

In a digital age, organizing and mining text can be an overwhelming task, especially when the amount of data is extensive. For example, comprehensive literature reviews of articles from 1970-2008 could easily include more than 1,000 articles exceeding 3,000,000 words for analysis. Data analysis of such "big literature" (Nunez-Mir, Iannone, Pijanowski, Kong, & Fei, 2016) can be resource intensive of time and people. Using effective and appropriate approaches for data organization and analysis is essential.

A number of software tools are available to analyze and organize qualitative data such as NVivo (http://www.qsrinternational.com/nvivo/what-is-nvivo), Atlas.ti (https://atlasti.com/), MAXQDA (https://www.maxqda.com/), and Dedoose (https://www.dedoose.com/). Concept mapping tools such as CmapTools (Cañas *et al.,* 2004) and Leximancer have also been employed in qualitative data analysis. These concept mapping tools allow for making meaning and connections, and synthesizing information and research data in a unique way. Most researchers tend to use a single tool for data analysis and do not consider how these tools can work in complementary ways to maximize data analysis. This paper presents the results of an exploratory comparative analysis case study using two concept mapping tools for textual analysis of large data. The two tools were used in combination but employed two approaches. This paper discusses the approaches implemented for using the two concept mapping tools in combination, the strengths and limitations of each approach, and suggestions for replicating the use of the two concept mapping tools in combination in other contexts.

2 Literature Review

Concept mapping has been recognized as an effective process for identifying concepts and drawing connections between the concepts (Baugh, McNallen, & Frazelle, 2014; Butler-Kisber & Poldoma, 2010). This functionality of concept maps enables users to analyze various forms of textual data and draw out main concepts and connections between them. Therefore, concept maps are often utilized to make meaning of participant interviews or to review bodies of literature. Conceição, Samuel, and Yelich Biniecki (2017) identified CmapTools and Leximancer as two concept mapping tools that were used in analyzing textual data.

2.1 CmapTools

CmapTools is based on Novak and Gowin's (1984) method of "organizing and representing knowledge" (p. 1). Concepts are represented in boxes and are connected to each other via lines that have linking words on them to show the relationship between the concepts. Concept maps are represented hierarchically with the most general concepts placed at the top of the map and becoming more detailed as the map develops. Conceição, Samuel, and Yelich Biniecki (2017) note that this relational approach to concept mapping has been utilized to conduct textual analysis and present research findings.

Relational approaches for using concept maps for data analysis have been used to identify themes in a literature review, summarize interview transcripts, and identify interconnectedness among concepts (Conceição, Samuel, & Yelich Biniecki, 2017). Vanderheide, Moss, and Lee (2013) used relational concept maps in the first phase of their literature review study. They created a concept map to identify the main concepts in "moral habitability," and they then used these concepts to guide their literature review. Kinchin, Streatfield, and Hay (2010) used concept maps "as a way of representing information gathered during research interviews" (p. 64). Baugh, McNallen, and Frazelle (2014) used relational concept maps to identify interconnectedness of data in historical research. While research studies have used CmapTools for organizing data, CmapTools is more often employed for presenting rather than analyzing data. Leximancer, on the other hand, has been primarily developed for the purpose of analyzing large data sets.

2.2 Leximancer

The Leximancer software uses an automated process where an algorithm mines the textual data and extracts concepts based on the frequency of word usage within the text and concepts are clustered based on their co-occurrence. Leximancer technology was created by Dr. Andrew E. Smith at The University of Queensland, Australia after seven years of research and development (Leximancer, n.d.). Smith and Humphreys (2006) explain that Leximancer,

> goes beyond keyword searching by discovering and extracting thesaurus-based concepts from the text data, with no requirement for a prior dictionary, although one can be used if desired. These concepts are then coded into the text, using the thesaurus as a classifier. The resulting asymmetric concept co-occurrence information is then used to generate a concept map. (p. 262)

Leximancer shows the results in the form of a concept map where each concept in the map is represented as a dot. The greater the frequency of the concept, the larger the dot. Concepts that co-occur are placed near each other; the distance between them indicating their frequency of co-occurrence. Martin and Rice (2007) and Pendergast, Garvis, and Kanasa (2011) discovered "congruence" in the themes identified by Leximancer and the researchers.

Smith and Humphreys (2006) define the map created by Leximancer as a concept map. Leximancer has been used extensively in research as a concept mapping tool (Liesch, Håkanson, McGaughey, Middleton, & Cretchley, 2011; Noble, O'Brien, Coombes, Shaw, & Nissen, 2011; Poser, Guenther, & Orlitzky, 2012). However, the map produced by Leximancer is vastly different from the Novakian model. Leximancer concept maps lack hierarchy and relationships between concepts. It is important to note that the term "concept map" is loosely used in the context of Leximancer.

Since Leximancer is a software tool, it can conduct content analysis on large data sets and generate concept maps much faster than is possible by humans. Researchers have, therefore, used Leximancer to analyze large data sets of blog posts and posts and comments on community and social media websites. Nunez-Mir et al. (2016) talk of the "big literature" (p. 1262) phenomenon, referring to the large amounts of scientific and academic literature that is being published. They contend that only automated content analysis through software like Leximancer can help "quantify and describe the existing literature on a topic in its entirety" (p. 1266). The speed of Leximancer enabled Liesch et al. (2011) to conduct a literature review of 1,249 articles spanning the years of 1970-2008 while Poser et al. (2012) reviewed 165 journal articles. Leximancer allows for a more rapid analysis of data and changes to the analysis can be repeated more efficiently; however, it does not mean that the overall process is faster as there is a need to clean and prepare the data for analysis.

While the rapid data analysis using Leximancer was an advantage, Cretchley et al. (2010) acknowledge that it lacks a nuanced understanding of themes and their relationships that researchers' repeated reading of texts provides. Noble et al. (2011) study highlights that a textual analysis can overlook key concepts, which then must be introduced manually. Kyle, Nissen, and Tett (2008) caution that Leximancer's reliability is higher with single-word concepts and reliability drops when multi-word concepts are included.

An analysis of the literature reveals that both tools, CmapTools and Leximancer, have strengths and limitations. While Leximancer provides speed of analysis, it overlooks relationships between key concepts and the nuances of

concepts. In addition, reliability is dependent on the words that compose each concept. While the creation of concept maps using CmapTools may be time consuming and require specialized training, the resulting analyses produce relationships that are more meaningful and relevant.

Text analysis studies, which incorporate concept maps, tend to utilize one or the other concept mapping tool. They do not use both tools within one study. Therefore, the purpose of this exploratory study was to investigate two approaches for data analysis and organization of large textual data that use the two concept mapping tools, CmapTools and Leximancer, in combination.

3 Methodology

Two implementation approaches were employed to compare the strengths and limitations of using CmapTools and Leximancer in combination. The research questions answered by this study are:

How can concept mapping tools be used to effectively and efficiently analyze and organize large textual data?

- What are the strengths of each approach?
- What are the limitations of each approach?
- How do the tools complement each other?

This research was conducted as an exploratory comparative analysis case study since the two concept mapping tools have not been used in combination to analyze data (Conceição, Samuel, & Yelich Biniecki, 2017). The lack of detailed preliminary research in this area inhibits the use of more complex research designs. This methodology allowed for examining two implementation cases to compare the strengths and limitations of the two approaches using concept mapping and to provide information for potential replication (Lewis-Beck, Bryman, & Futin, 2004).

3.1 Context of Study

Two implementation cases were employed in the context of a large publication containing four volumes in the field of adult and continuing education. The four-volume publication aimed at creating a major reference work covering the knowledge base of the field. The editors of the publication recognized that effective and appropriate processes were needed to help organize themes within each volume, identify relationships among themes within a volume, and conceptually link different articles within the overall publication. The intent was to identify concepts in a more global way and ensure consistency of concepts across the different articles within volumes.

Written by practitioners and scholars from different countries and representing diverse parts of the field, the publication contains 80 articles ranging from 200 to 3,000 words each spread over the four volumes. The study focused on the approaches employed by two volumes within the publication. The two concept mapping tools were initially implemented within Volume 2 of the publication. To explore the effectiveness and appropriateness of the two tools, the implementation order was then inverted and applied within Volume 3 of the publication.

The “approach” is the order of implementation of tools while the “case” refers to the operationalization of the approaches. In the implementation of Case 1, Leximancer was used first to conduct a broad analysis of the large textual data and draw out the main concepts. Then CmapTools was employed to identify themes and relationships among articles in Volume 2. The process was inverted in the implementation of Case 2 where CmapTools was employed first followed by Leximancer. These implementation cases allowed for the exploration of the strengths and limitations of the two tools and how they worked best in combination with each other.

3.2 Implementation of Case 1 (Approach 1)

Volume 2 of the publication used the concept mapping approaches in the following order:

1. Leximancer was used to identify the frequency of words and visually present the co-occurrence of concepts and the distance between the concepts. To do this, reference lists, author names, keywords, and abstracts were removed from all articles in Volume 2.
2. The cleaned data were uploaded into Leximancer and the analysis was run. Leximancer generated a list of themes and the number of hits based on co-occurrence. These themes are shown from most relevant to least

relevant in Table 1, along with concepts close to these themes and the number of hits. The file generated by Leximancer also provided the sentences where these themes were located, which are not included in this paper. In Leximancer, a concept is a group of related words or terms that are located within proximity in the text.

3. Then, the concepts were viewed through a concept map generated by Leximancer. Figure 1 shows two formats for presenting the major themes that had the most hits and the location of concepts near them. The dots and lines connecting the dots show connections among the themes and concepts. For example, "learning" re-occurred the most; it appears in red and is connected to other concepts of "experience," "social," "development," "professional," "experiences," "practice, "knowledge," and "personal." Colored-themes on a Leximancer concept map have a specific meaning. For instance, hot colors (such as red and orange) indicate the most important themes while cool colors (such as blue and green) imply less relevance.

Theme	Concepts	Hits
Learning	learning, social, development, professional, experiences, practice, knowledge, personal	308
Adult	adult, education, educators, cultural	205
Learners	learners, programs, critical	143
Relationships	relationships, mentoring, protégés, mentors, experience, theory	106
Provide	provide, educational	65
Community	Community	43
Culture	culture, popular	40
Program	Program	38
Power	Power	29
Need	Need	28
Skills	skills	21
Example	example	20
model	model	20

Table 1: Leximancer List of Major Themes, Concepts, and Hits

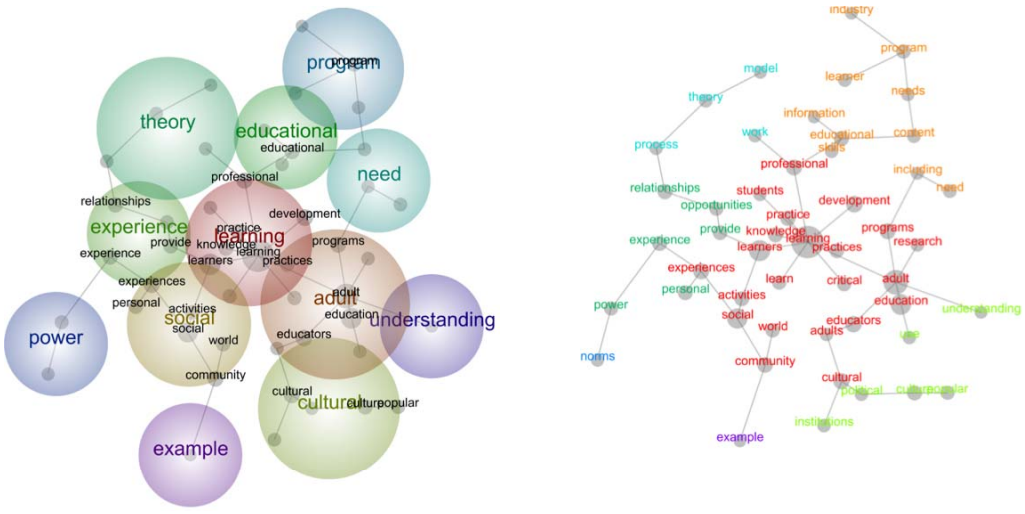

Figure 1. Leximancer generated concept maps in Case 1.

4. Since the concepts with lower relevance were too broad in the context of the articles, the specific articles where these terms occurred were extracted and analyzed separately. For example, the themes of "community," "culture," "program," "power," "need," and "skills" had 20 to 40 hits but did not have other concepts near them; however, they were of great relevance in the context of the publication as seen by the interconnections shown in Figure 1 through dots and lines. The theme of "example" was a common word used by article authors based on the publication guidelines. Articles should provide practical examples of

adult education; thus, the word "example" was often included in the articles. The word "example" was not considered relevant. In this case, Leximancer generated concepts allowed to narrow their scope.

5. The volume editors, who had read all the articles for Volume 2, reviewed the concept ranking from Leximancer and the key concepts extracted, and developed the table of contents for Volume 2.
6. Based on the table of contents and clues from Leximancer analysis, the introduction for the volume was developed.
7. Using the volume introduction, a concept map employing CmapTools was drawn out to identify relationships between concepts among articles. The key concepts from the table of contents were extracted and identified as the main concepts for Volume 2. Suggestions were made for the inclusion of other concepts. Figure 2 shows the volume concepts and their interrelationships using CmapTools. The CmapTools version of the volume provides a graphical representation of the content in the volume as a way to market the publication and provide a visual summary of the content.

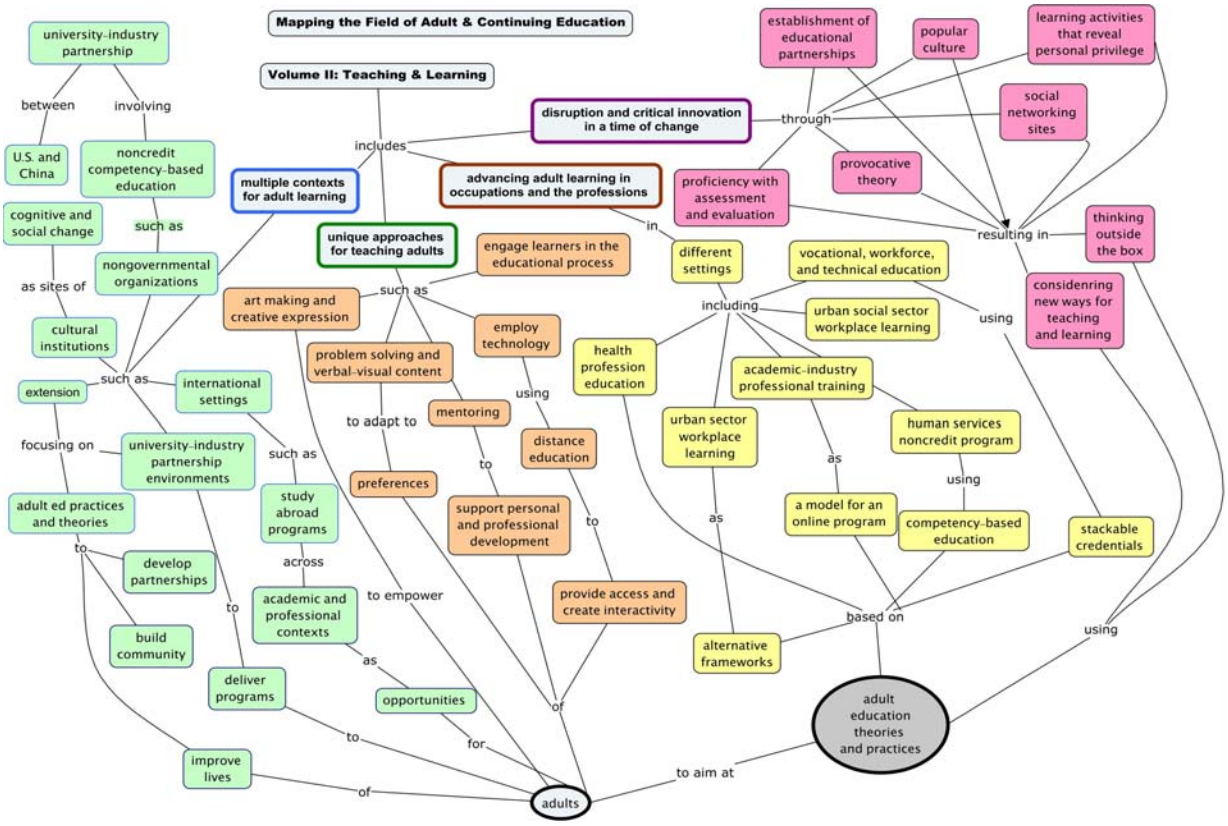

Figure 2. Concept map using CmapTools in Case 1.

3.3 Implementation of Case 2 (Approach 2)

In Volume 3 of the publication, concept mapping approaches were implemented retroactively after the volume had been published. Volume 3 of the publication used the concept mapping approaches in the following order:

1. This approach started with the creation of a relational concept map using CmapTools. The map was developed using the article keywords, table of contents in the volume, the introduction to the volume, and key ideas identified by the consulting editors. (See Figure 3 for the concept map using CmapTools in Case 2)
2. Then, Leximancer content analysis was conducted on all the articles in Volume 3. To enable this, references, abstracts, keywords, and author names were removed from the articles. These clean data were

then run through Leximancer and a concept map was generated. (See Figure 4 for the Leximancer generated concept map in Case 2)

3. The concept map and concept ranking developed by Leximancer were reviewed against the concept map created with CmapTools to provide triangulation of data. At this point, it was observed that the concepts of "time," "needs," "organization," and "social," that were ranked highly by Leximancer, were not represented in the concept map created using CmapTools.
4. The editors drilled down into the occurrences of these concepts, as identified by Leximancer, and discovered that concepts of "time" and "needs" lacked nuanced analysis. Hence, though ranked highly in frequency, they were not meaningful in the overall scope of all the articles. The concept of "social" was repeated often in just two articles but it did not capture the overall essence of the volume. However, re-examining the concept map created with CmapTools led to the addition of "collective advocacy" as a better representation of the concept of "social." Re-examination of this concept map also revealed that the theme of "organization" was embedded in the meaning structure of the map and appeared under different nomenclature such as "organizational culture." At the same time, it was noted that the addition of the concepts of "adult administrative leaders" and "social leaders" needed to be included.

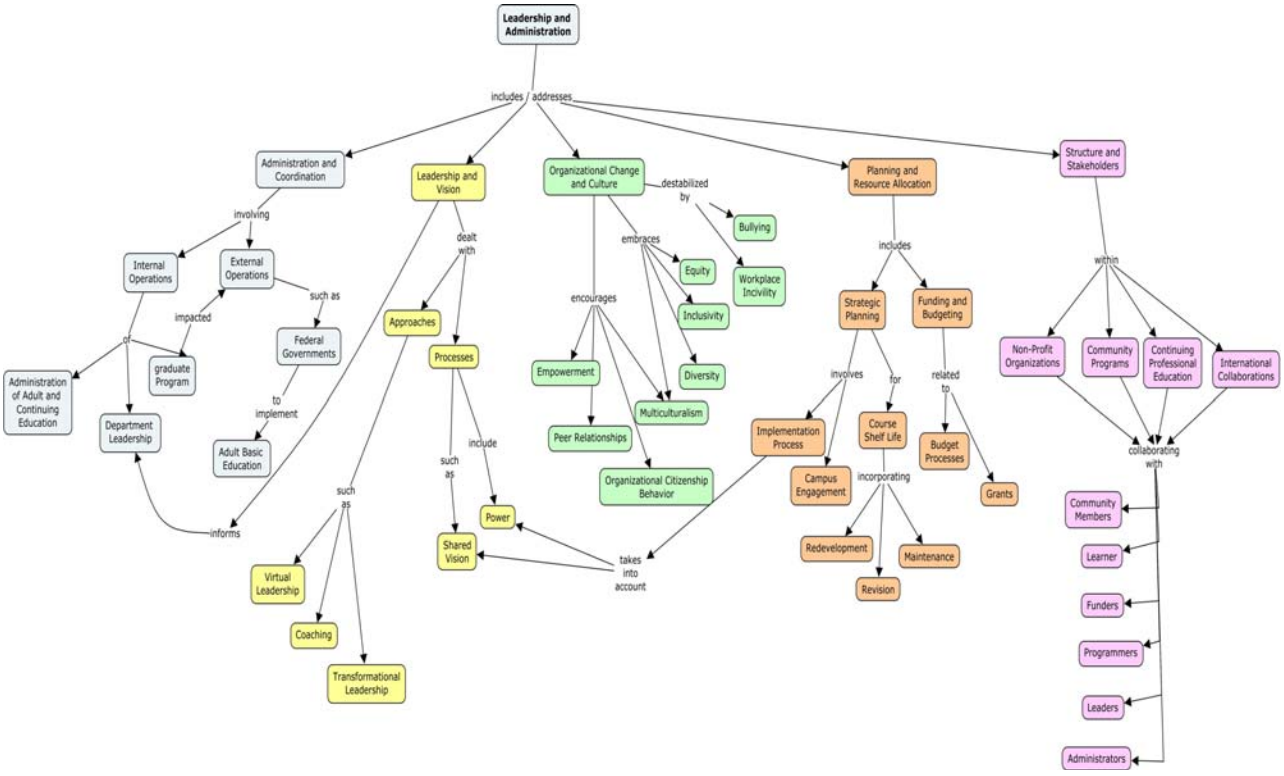

Figure 3. Concept map using CmapTools in Case 2.

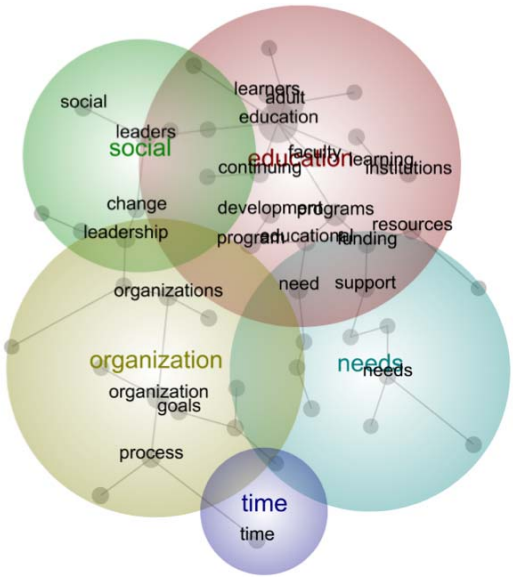

Figure 4. Leximancer developed concept map in Case 2.

5. In addition to concepts, the distances between concepts identified by Leximancer led to the development of additional cross-links in the concept map created with CmapTools. This concept map was refined and appropriate and meaningful changes were made. (See Figure 5 for the revised concept map using CmapTools for Case 2)

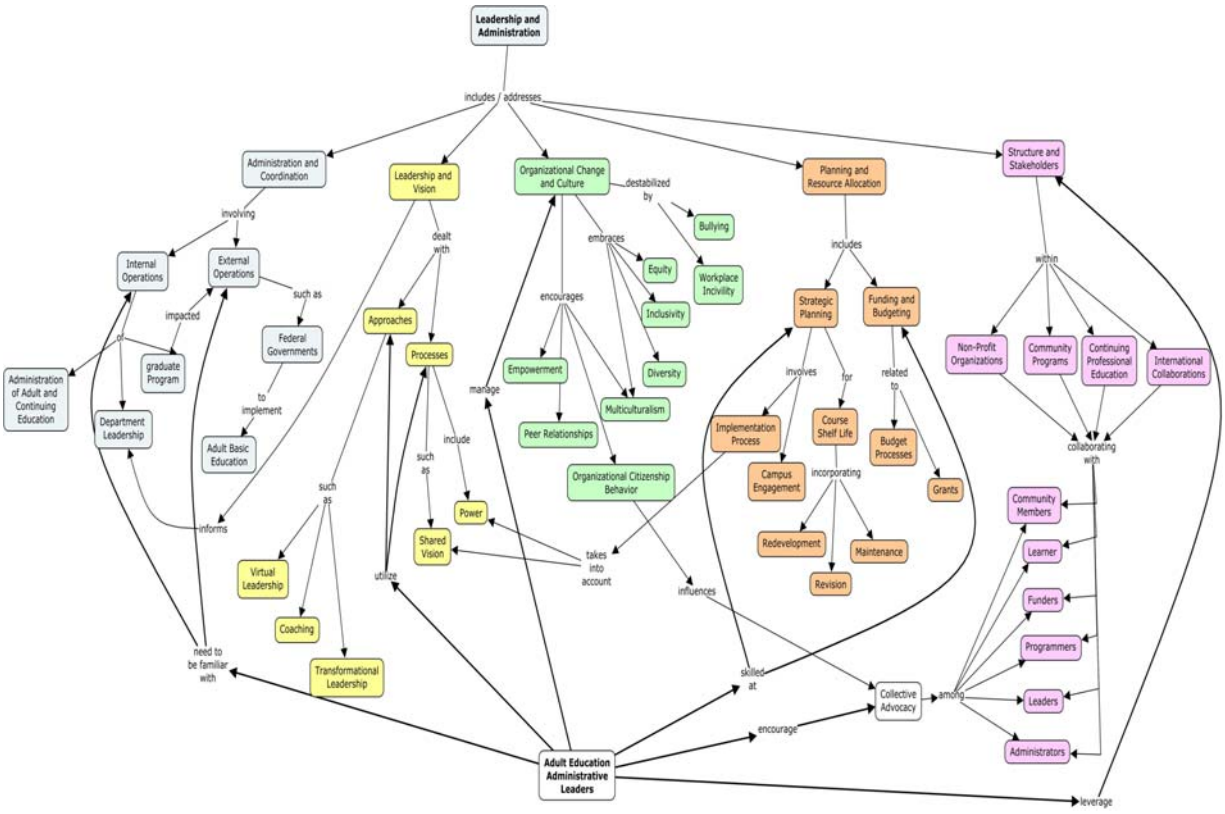

Figure 5. Revised concept map using CmapTools in Case 2.

4 Results

This exploratory study compared two approaches to data analysis and organization using CmapTools and Leximancer in combination. Both approaches revealed that CmapTools and Leximancer, used individually, left gaps in the data analysis. While Leximancer could draw out themes and relationships quickly, the textual data needed to be scrubbed carefully to avoid inaccuracies. For example, including references in the data set led to the identification of a repeated concept that was irrelevant such as the name of a city in a reference list. In order for words to be contextualized, terms needed to be defined accurately for the Leximancer library. This is a time-consuming process especially for beginners. It takes several iterations to refine the concepts until the researcher is proficient with Leximancer.

Meagher-Stewart et al. (2012) note that the relational approach to concept mapping is limited in the amount of time involved to conduct analysis since the researcher is often the tool of analysis. The researcher is solely responsible for creating the concept map including identifying the concepts and structuring a hierarchy, defining propositions, and relating concepts through cross-links. Familiarity with the software is required in order to use them effectively and appropriately. As a software program, CmapTools also involves a learning curve. It requires some form of training in order to create maps with hierarchical and propositional relationships.

Both tools were time consuming in different ways. Although using the concept mapping tools individually highlighted limitations, using CmapTools and Leximancer in combination reduced the limitations of each tool making both approaches effective. Both approaches provided ways to analyze and organize large textual data in their own ways. When using both concept mapping tools together, they complemented each other and provided distinctive outcomes.

In approach 1, using Leximancer first allowed for about 25,000 words from one volume to be input into the software program to draw out and identify relevance of the concepts within a few minutes. The use of the software program enabled effective identification of broad themes facilitating subsequent collaborative brainstorming and the creation of the table of contents for the publication. The deep dive into the text using the CmapTools second enabled the meaningful connection of appropriate concepts across articles in Volume 2. The relational concept map can give intended readers a graphical summary of the content in the large publication as a way to select the more relevant articles to them by viewing the interrelationships among concepts before delving into the article.

In approach 2, using CmapTools first followed by Leximancer allowed for triangulation of data and helped identify missing themes and insightful relationships more effectively and efficiently. The process gave a global view of Volume 3 with a more insightful perspective.

5 Conclusion

The availability of concept mapping tools such as CmapTools and Leximancer make concept mapping more accessible to researchers. The use of the two concept mapping tools in combination affords a unique opportunity to enrich the process of data analysis. The effectiveness of the approaches is in balancing the strengths and limitations of each concept mapping tool and leveraging their functionality based on the purpose of their use.

Replication of the two approaches is best dictated by the data and research teams. For example, when working with big literature, such as a literature review of journal articles published on a topic over the past 20 years, it is more effective to use approach 1 and begin analysis using Leximancer. Broad themes and connections can be identified, and researchers can then delve into relevant sections to clarify the relationships between the concepts. When analyzing data in research teams, using Leximancer first ensures a consistency in the main concepts identified. Smaller data sets, like interviews with a few participants conducted by one researcher, might benefit from approach 2. Researchers can analyze the data using CmapTools and use Leximancer for triangulation of data in order to confirm, expand, or broaden data analysis, interpretation, and presentation.

One serious consideration with the use of Leximancer is the cost of the software tool itself. While CmapTools is open-access, Leximancer does have licensing costs associated with the software. Researchers should consider their

return on investment. If Leximancer is to be used on an ongoing basis, the cost might be justified. However, for a stand-alone project, Leximancer might prove to be too expensive.

This paper is not an attempt to prove that one tool is better than the other. Rather, the aim of this study was to explore how research can be strengthened when multiple tools are used in combination. This study assessed two ways of combining two concept mapping tools, CmapTools and Leximancer, to see if they would benefit research and data analysis. The findings of this study show that while each tool can be used individually, they are more powerful when used in combination.

References

Baugh, N., McNallen, A., & Frazelle, M. (2014). Concept Mapping as a Data Collection and Analysis Tool in Historical Research. *Qualitative Report*, *19*(13), 1-10.

Butler-Kisber, L., & Poldma, T. (2010). The Power of Visual Approaches in Qualitative Inquiry: The Use of Collage Making and Concept mapping in Experiential Research. *Journal of Research Practice*, *6*(2), 1-17.

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Lott, J, Carvajal, R. (2004). CmapTools: A Knowledge Modeling and Sharing Environment. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology*. Proc. of the First Int. Conference on Concept Mapping (Vol. I, pp. 125-133). Pamplona, Spain: Universidad Pública de Navarra.

Conceição, S. C., Samuel, A., & Yelich Biniecki, S. M. (2017). Using Concept Mapping as a Tool for Conducting Research: An Analysis of Three Approaches. *Cogent Social Sciences*, *3*(1), 1404753.

Cretchley, J., Gallois, C., Chenery, H., & Smith, A. (2010). Conversations Between Carers and People with Schizophrenia: a Qualitative Analysis using Leximancer. *Qualitative Health Research*, *20*(12), 1611-1628.

Kinchin, I. M., Streatfield, D., & Hay, D. B. (2010). Using Concept Mapping to Enhance the Research Interview. *International Journal of Qualitative Methods*, *9*(1), 52-68.

Kyle, G. J., Nissen, L., & Tett, S. (2008). Perception of Prescription Medicine Sample Packs among Australian Professional, Government, Industry, and Consumer Organisations, based on Automated Textual Analysis of One-on-One Interviews. *Clinical Therapeutics*, *30*(12), 2461-2473.

Lewis-Beck, M. S., Bryman, A. & Futing Liao, T. (2004). *The SAGE Encyclopedia of Social Science Research Methods.* Thousand Oaks, CA: SAGE Publications Ltd. doi: 10.4135/9781412950589

Leximancer. (n.d.). About Leximancer. Retrieved from: https://info.leximancer.com/company/

Liesch, P. W., Håkanson, L., McGaughey, S. L., Middleton, S., & Cretchley, J. (2011). The Evolution of the International Business Field: A Scientometric Investigation of Articles Published in its Premier Journal. *Scientometrics*, *88*(1), 17-42.

Martin, N. J., & Rice, J. L. (2007). Profiling Enterprise Risks in Large Computer Companies using the Leximancer Software Tool. *Risk Management*, *9*(3), 188-206.

Meagher-Stewart, D., Solberg, S. M., Warner, G., MacDonald, J., McPherson, C., & Seaman, P. (2012). Understanding the Role of Communities of Practice in Evidence-Informed Decision Making in Public Health. *Qualitative Health Research, 20*(10),1-17.

Noble, C., O'Brien, M., Coombes, I., Shaw, P. N., & Nissen, L. (2011). Concept Mapping to Evaluate an Undergraduate Pharmacy Curriculum. *American Journal of Pharmaceutical Education*, *75*(3), 55: 1-11.

Novak, J.D., & Gowin, D. B. (1984). Learning How to Learn. Cambridge: Cambridge University Press.

Novak, J. D., & Cañas, A. J. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

Nunez-Mir, G., Iannone, B., Pijanowski, B., Kong, N., & Fei, S. (2016). Automated Content Analysis: Addressing the Big Literature Challenge in Ecology and Evolution. *Methods in Ecology and Evolution, 7*(11), 1262-1272.

Pendergast, D., Garvis, S., & Kanasa, H. (2011). Insight from the Public on Home Economics and Formal Food Literacy. *Family and Consumer Sciences Research Journal*, *39*(4), 415-430.

Poser, C., Guenther, E., & Orlitzky, M. (2012). Shades of Green: Using Computer-Aided Qualitative Data Analysis to Explore Different Aspects of Corporate Environmental Performance. *Journal of Management Control, 22*(4), 413- 450.

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior research methods, 38*(2), 262-279.

Vanderheide, R., Moss, C., & Lee, S. (2013). Understanding Moral Habitability: A Framework to Enhance the Quality of the Clinical Environment as a workplace. *Contemporary Nurse, 45*(1), 101-113.