Identifying Job Categories and Required Competencies for Instructional Technologist: A Text Mining and Content Analysis

Le Chen

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Curriculum and Instruction

Kenneth R. Potter, Chair
Barbara B. Lockee
Katherine S. Cennamo
Mark A. Bond

May 6, 2020
Blacksburg, Virginia

Keywords: text mining, content analysis, job analysis, competency, T-LAB, topic modeling

Identifying Job Categories and Required Competencies for Instructional Technologist:

A Text Mining and Content Analysis

Le Chen

ABSTRACT

This study applied both human-based and computer-based techniques to conduct a job analysis in the field of instructional technology. The primary research focus of the job analysis was to examine the efficacy of text mining by comparing text mining results with content analysis results. This agenda was fulfilled by using job announcement data as an example to determine essential job categories and required competencies. In phase one, a job title analysis was conducted. Different categorizing strategies were explored, and primary job categories were reported. In phase two, the human-based content analysis was conducted, which identified 20 competencies in the knowledge domain, 22 in the ability domain, 23 in the skill domain, and 13 other competencies. In phase three, text mining (topic modeling) was applied to the entire data set, resulting in 50 themes. From these 50 themes, the researcher selected 20 themes that were most relevant to instructional technology competencies. The findings of the two research techniques differ in terms of granularity, comprehensibility, and objectivity. Based on evidence revealed in the current study, the author recommends that future studies explore ways to combine the two techniques to complement one another.

Identifying Job Categories and Required Competencies for Instructional Technologist: A Text Mining and Content Analysis

Le Chen

GENERAL AUDIENCE ABSTRACT

According to Kimmons and Veletsianos (2018), text mining has not been widely applied in the field of instructional technology. This study provides an example of using text mining techniques to discover a set of required job competencies. It can be helpful to researchers unfamiliar with text mining methodology, allowing them to understand its potentials and limitations better. The primary research focus was to examine the efficacy of text mining by comparing text mining results with content analysis results. Both content analysis and text mining procedures were applied to the same data set to extract job competencies. Similarities and differences between the results were compared, and the pros and cons of each methodology were discussed.

Acknowledgments

First, I would like to express my deepest gratitude and appreciation to my advisor, Dr. Ken Potter, for spending countless hours with me throughout my graduate studies and dissertation research. His words and guidance brought me to where I am today. Thank you, Dr. Potter, for always being patient and supportive.

Many thanks also go to other committee members, Dr. Barbara Lockee, Dr. Katherine Cennamo, and Dr. Aaron Bond. All of them have supported me and provided valuable feedback along the way.

Dr. Cennamo also offered me the opportunity to visit Virginia Tech in the first place. This unique opportunity made me decide to start a new chapter in my life. I will be forever grateful for her kindness.

I would also like to thank Dr. Qing Zhang and Mingyu Li, who assisted me in conducting the content analysis part of my research and offered me valuable suggestions.

Special thanks also go to Dr. Glen Holmes for constantly encouraging me during my final year working as his graduate assistant.

Last but not least, I want to thank my family (my parents, wife and son) for allowing me to focus on the school work. Without their continuous support, the completion of this degree would not be possible.

Table of Contents

List of Figures

List of Tables

Chapter One - Introduction

Problem Statement

Throughout the years, competence-based approaches have been adopted as an essential framework in training and education programs (Sampson & Fytros, 2008). Thus, numerous efforts have been made to examine the professional competencies needed for instructional technologists. The term instructional technologist refers to, "a person who is employing the instructional development process to solve learning and performance problems and needs in a technology-based learning environment" (Tennyson, 2001, p. 356). Traditionally, the fields of educational technology, instructional technology, and instructional design and technology were used interchangeably in literature. For the purpose of simplicity and consistency, this paper refers to the field as instructional technology, and its practitioners as instructional technologists.

Professional associations such as the Association of Educational and Communication Technology (AECT), the International Society for Performance Improvement (ISPI), the Association for Talent Development (ATD), and the International Society for Technology in Education (ISTE) have all released different standards, guidelines, or frameworks for practitioners in the field.

Apart from that, many scholars have conducted job analyses to reveal competencies for instructional technologists with various foci (Byun, 2000; Kang & Ritzhaupt, 2015; Moallem, 1995; Ritzhaupt, Martin, & Daniels, 2010; Ritzhaupt, Martin, Pastore, & Kang, 2018; Shank, 2006; Sugar, Hoard, Brown, & Daniels, 2012). Some researchers have examined professional competencies in different work environments (such as business, industry, government, military, university, college and school district, etc.) (Byun, 2000; Kang & Ritzhaupt, 2015; Moallem, 1995; Morlan & Lu, 1993; Ritzhaupt et al., 2018). Others have focused on one context such as higher education (Marcial & Jeambe, 2014; Ritzhaupt & Kumar, 2015; Surry & Robinson, 2001), or one specific role such as projector manager (Brill, Bishop, & Walker, 2006), instructional design librarian (Osorio, 1999; Shank, 2006), or multimedia specialist (Ritzhaupt et al., 2010; Sugar, Brown, Cafeteria, & Daniels, 2007; Sugar et al., 2012).

These collective efforts have made a significant contribution to the field by informing theory and practice. This tradition of competency inquiry remains relevant and will most likely be a long-lasting endeavor. This is partially because the field is constantly growing, and partially due to the emergence of new technologies such as social media, mobile technology, artificial intelligence, etc. As a result, new studies are always needed to keep the knowledge base up-to-date.

The field of instructional technology is multidisciplinary in nature, and incorporates a wide range of subdomains (Ritzhaupt et al., 2010). Some have been extensively investigated, while others have not. For example, in higher education, Surry and Robinson (2001) analyzed 449 job announcements, and presented a taxonomy of instructional technology service positions in higher education consisting of eight job types: (a) instructional technologist, (b) instructional designer, (c) distance learning coordinator, (d) instructional technology manager/administrator, (e) technical support specialist, (f) Web specialist, (g) instructional technology librarian, and (h) miscellaneous. However, in K-12 education, due to a lack of studies, it is unclear whether the same job categories are needed. Some researchers have analyzed instructional technology competencies needed for school teachers and administrators, but not for instructional technologists working in K-12 schools (Lu & Miller, 2002; Northrup, 1997). Byun (2000) recommended that future research should gather data from professional and commercial job banks and focus on specific job categories to increase the accuracy of the competencies.

The final reason for proposing a new study that follows the tradition of competency inquiries is due to the availability of new research techniques applicable to job announcement analyses. In recent years, more and more job announcement studies were conducted using algorithm-based approaches such as text mining (De Mauro, Greco, Grimaldi, & Ritala, 2018; Maer-Matei, Mocanu, Zamfir, & Georgescu, 2019; Uhm, Lee, & Jeon, 2017; Wowczko, 2015; Yang et al., 2016). Kimmons and Veletsianos stated that mining of public Internet data offers huge benefits, such as, "providing larger amount of data and allowing easy randomizing, empowering both quantitative and qualitative analyses, enabling identification of subgroups for further research, and avoiding many biases" (2018, p. 493). According to them, job advertisement is a data source compatible with text mining techniques.

Research Questions

This study aims to answer the following research questions:

1. How might job categories be identified when using multiple approaches to aggregating job titles?

Instructional technologists work in a variety of contexts with various job titles. Identifying job categories can be helpful in understanding career opportunities for instructional technologists. Since the literature is inconclusive about the best technique to categorize job titles, this study explores multiple approaches to see how job categories can be clustered from different perspectives.

2. What instructional technology competencies might be identified from human-based content analysis?

This question is designed to elicit findings from human-based analysis - allowing human coders to extract competency statements from job descriptions based on a given theoretical framework.

3. What instructional technology competencies might be identified from computer-based text mining?

This question aims to explore the capability of text mining to see how job competencies can be identified with minimal human intervention.

4. How do the results of text mining differ from human-based content analysis?

The answer to this question will be based on the findings from research questions two and three. Similarities and differences will be discussed. The focus is to compare the differences in terms of granularity, comprehensibility, and objectivity.

Purpose of the Study

This study applies both human-based and computer-based techniques to job announcement data in the field of instructional technology. The primary research focus was to examine the efficacy of text mining by comparing the results with content analysis. This agenda was fulfilled in two steps: (a) applying text mining in the field of instructional technology using job analysis as an example to determine essential job categories and required competencies, and (b) comparing the findings of text mining with content analysis in order to examine the pros and cons of each method. By comparing text mining results with human-based content analysis, we can have a better understanding whether the two techniques are compatible from the methodology perspective.

Significance of the Study

This study adds value to the field in several ways: it contributes to the knowledge base by identifying recent job categories and competencies, which might be informative for practitioners working in K12 and higher education. However, the main contribution is the exemplary research methodology showing how text mining can be applied to job analysis.

According to Kimmons and Veletsianos (2018), text mining has not been widely applied in the field of instructional technology. As a field, we have not fully investigated the potentials and ramifications of this emerging research methodology. This study provides an example of using text mining techniques to discover required job competencies. This may be helpful to researchers unfamiliar with the methodology, allowing them to better understand its potentials and limitations.

Chapter Two - Review of Literature

Conceptual Framework

According to Merriam Webster Online (2020), competence and competency are identical words meaning, "a sufficiency of means for the necessities and convenience of life." They also mean, "possession of sufficient knowledge or skill."

Historically, the term "competency" was used in different contexts and with a slightly different meaning (Bullough Sr. & Brumbaugh, 1974). Researchers acknowledged this ambiguity in literature, and efforts were made to refine the definition for clarity and better communication. For instance, Gale and Pol (1975, p. 21) searched the meanings in multiple dictionaries and found considerable agreement across the sources. They ultimately defined the term competence as, "the quality of being functionally adequate in performing the tasks and assuming the role of a specified position." They stated that the defining feature of competencies is that they are associated with job requirements and responsibilities. Competencies can be seen as a combination of skill, knowledge, and attitude. Those who have such capabilities are considered good candidates for the position.

It is important to acknowledge that researchers differ in their understanding and definition of what competency is. In general, two approaches can be found in literature. The first approach identifies competency as behavioral goals, which are measurable through observable performance outcomes such as:

- "Measurable human capabilities that are required for effective work performance demands" (Marrelli, 1998, p. 8)
- "Distinguishable elements of underlying capacities or potentials which allow job incumbents to act competently in certain situations" (Ley & Albert, 2003, p. 2)

- “Competency refers to an individual's ability to perform a specific task or deliver a measurable outcome” (Succar, Sher, & Williams, 2013, p. 2)

The second approach views competency as a combination of knowledge, skill, and personal traits that can be used as a measure to predict future performance (Succar et al., 2013). Examples are:

- “A knowledge, skill, or attitude that enables one to effectively perform the activities of a given occupation or function to the standards expected in employment.” (Richey, Fields, & Foxon, 2001, p. 31)
- “A cluster of related knowledge, skills and attitudes (KSA) that affects a major part of one's job (a role or responsibility), that correlates with performance on the job, and that can be improved via training and development” (Parry, 1996, p. 50)
- “A practical ability and dexterity; knowledge; understanding; ability; proficiency” (Stolovitch, Keeps, & Rodrigue, 1995, p. 44)

In the field of instructional technology, the second approach seems to be in favor. For example, Richey et al. (2001, p. 31) defined competency as, “knowledge, skill or ability that enables one to effectively perform the activities of a given occupation or function to the standards expected in employment.” This definition aligns with the one specified by IBSTPI (2017) as, “an integrated set of skills, knowledge, and attitudes that enables one to effectively perform the activities of a given occupation or function to the standards expected” (p. 1).

Even though the term competency has different meanings, as outlined above, for the purposes of this study, competency will be used to refer to a combination of knowledge, skills, abilities, and other characteristics needed to perform a specific task (KSAO, including other characteristics). This approach was selected for the current study because KSAO is the most commonly used framework around the world (Uhm et al., 2017), and has been widely adopted and rigorously validated by many researchers in the field of instructional technology (Durmaz, 2012; Ritzhaupt et al., 2010; Ritzhaupt & Kumar, 2015).

Competency Standards

Professional associations have profoundly influenced the field of instructional technology. One valuable contribution is by defining the standards of professional competency for the field. These associations include the Association for the Advancement of Computing in Education (AACE), the Association of Educational and Communication Technology (AECT), the American Society for Training and Development (ASTD), the International Board of Standards for Training Performance and Instruction (IBSTPI), the International Society for Performance Improvement (ISPI), and the International Society for Technology in Education (ISTE).

The AECT 2012 Standards were the product of a five-year development process. Under the framework of their 2008 definition, "Educational technology is the study and ethical practice of facilitating learning and improving performance by creating, using, and managing appropriate technological processes and resources" (Januszewski & Molenda, 2007, p. 1). The AECT (2012) standards highlight the professional competencies a candidate needs to possess in five domain areas, i.e., content knowledge, content pedagogy, learning environment, professional knowledge and skills, and research. Under each domain, descriptive indicators are provided. These standards have been widely adopted by many academic programs as curriculum guidelines. Though AECT is not an official accrediting body, its standards remain one of the most rigorous sets of guidelines for professionals in the field.

The ISTE standards are a framework that helps users to rethink the process of learning and to redesign innovative learning environments. To date, ISTE (2019) has released a collection of standards for different groups (students, educators, education leaders, coaches, and computer science educators). Additionally, there is a computational thinking standard for helping educators to integrate computational thinking into teaching practices. These standards can be applied to both formal and informal settings in education. ISTE has claimed that their standards have received recognition both nationally and internationally. Huggins, Ritzhaupt, and Dawson (2014) have also claimed that the ISTE standards could be used as reference tools for creating measurement instruments.

The IBSTPI competency model is a hierarchical framework containing three main components—domains, competencies, and performance statements (Spector, 2015). Competencies are considered to be the essential component, and each competency is described by a list of performance statements. The 2017 standards of Instructional Designer competency contain five domains: (a) professional foundation, (b) planning and analysis, (c) design and development, (d) evaluation and implementation, and (e) management, as well as 105 performance statements (IBSTPI, 2017). IBSTPI also provides standards for online learners, training managers, evaluators, and instructors. According to the IBSTPI, their performance statements were developed using a rigorous process, and validated with hundreds of researchers and practitioners from different work environments and geographical locations (Ritzhaupt et al., 2018).

In 2014, ATD, as a professional organization for training and development, updated its competency model defining the knowledge and skills required for talent development professionals. This model contained foundational competencies for everyone in the field as well as specific areas of expertise (AOEs) needed for specific roles. The 2017 version included ten AOEs based on the KSA framework. ATD recognized candidates who passed tests on all AOEs as Certified Professionals in Learning and Performance.

Job Analysis Approach

One approach to identifying competencies is to conduct surveys of professionals and experts, collecting their opinions on what is required to be an instructional technologist (Molenda & Pershing, 1992; Patterson, 1985; Spitzer, 1988). The primary disadvantage of the survey approach is its subjectivity, as the respondents take on the role of a researcher to give their interpretation of reality (May, 1980). Alternatively, a second approach, job analysis, seems to be less subjective. Job analysis is, "the process of identifying work activities and worker requirements or Knowledge, Abilities, Skills and Other Characteristics (KASOCs) of a set of positions sharing the same job title" (Sanchez, 1994, p. 1) .

Job announcements are developed and reviewed by many people in the hiring organization to ensure the accuracy of job responsibilities and job requirements. In many cases, department leaders, senior-level administrators, search committees, or recruiters are all asked to provide input in defining necessary competencies. It has been reported that, once a requirement is posted in the job announcement, it becomes difficult to hire a less qualified person (Byun, 2000). Thus, the wording for job announcements is usually reviewed carefully to avoid confusion or difficulty. Many researchers prefer a job analysis approach to reveal job competencies.

Mullins (1985) considered job analysis to be an important practice, stating that job analysis data could be used for, "performance review and appraisal, training, reward systems, staff development and career progression, etc."(p. 183). Singh (2008) described it as, "a sound business practice that can improve communication, accommodate change" (Singh, 2008, p. 89).

Researchers believe that job analysis studies can be helpful in many ways including: training and development, performance evaluation, compensation and promotion, job description, and job design (Mullins, 1985; Sanchez, 1994; Singh, 2008).

As a result, many job analysis studies have emerged within the last twenty years, including the field of instructional technology. For example, Moallem (1995) conducted a study that reported the job responsibilities of instructional technologists by analyzing 150 job announcements within a three-year period. The results of this study were summarized with a list of competencies for three different work environments (business and industry, government and military, and education). Osorio (1999) analyzed 201 Higher-Ed library positions in 1976, 1986, and 1998 to investigate how job responsibilities and requirements had changed over the years. The author stated that job announcement data collected from those three years could indicate how technology had impacted the role of librarians in the last decades. Shank (2006) analyzed job announcements for the newly emerging Instructional Design Librarian positions from 1999-2004. Ten unique position announcements were identified and examined to report job qualifications and responsibilities.

Content Analysis

With regard to research techniques, the traditional method of conducting a job analysis is via content analysis. According to Krippendorff (2004), content analysis is considered an essential research technique in social science, which is commonly used to discover knowledge and interpret meaning from textual and other type of data forms. Hsieh and Shannon (2005, p. 1278) defines it as, "a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns."

Weber (1990) states that content analysis can enable researchers to extract useful information from existing documents to make valid inferences on the subject of interest (Weber, 1990). Fraenkel, Wallen, and Hyun (2011) pointed out that content analysis, "enables researchers to study human behavior in an indirect way, through an analysis of their communications" (p. 476).

Content analysis can be used in both qualitative and quantitative inquiry (Weber, 1990). In quantitative content analysis, researchers tend to choose the deductive approach and use a theoretical framework to develop coding schemes (Neuendorf, 2010); in qualitative content analysis, researchers typically generate coding schemes inductively during the process of data analysis (Drisko & Maschi, 2015).

Typically, the first step is to convert other forms of data to text. Researchers then code textual data with tags. Tags can be described as, "categories at various levels, such as word, phrase, sentence, paragraph, or theme" (Gaur & Kumar, 2018, p. 281). The coding category is selected or created to represent the characteristics of interest according to research objectives. These categories are also called coding schemes, rules, or codebook. The coding scheme is applied to the entire data set to maintain the consistency of retrieved information. The researcher can use the newly-generated information to draw inferences on its own, or combine it with other data for further analysis.

As for coding techniques, Short and Palmer (2008) classified them into three categories: human-based system, word-count system, and artificial intelligence system. The human-based system is a manual coding process conducted by trained coders. Researchers make the decision about coding units for classification, create the coding scheme, which can be either deductive or inductive according to research objectives, and prepare a guideline or coding scheme with a description of each category. The guideline or coding scheme is for coder training purposes. The common understanding and content familiarity among coders are key to achieving a high level of inter-coder reliability.

For example, Barison and Santos (2011) conducted a qualitative content analysis of Building Information Modelling (BIM) job ads to identify which competencies a BIM manager should have. Gathercole and Thurairajah (2014) performed a quantitative content analysis of over 300 job ads to reveal BIM-targeted positions and required competencies. The coding process was performed manually by human coders in both studies.

Byun (2000) examined instructional technology competencies by analyzing job postings over a period of five years. A total of 827 job postings were analyzed, including 367 in corporate settings, 413 in education, and 47 in not-for-profit organizations. The researcher used an inductive procedure to analyze data by sorting and sifting categories throughout the coding phase. Any emerging category was recorded. After coding one-hundred job postings, results were examined, and overlapping competencies were merged.

Text Mining

Over the last few decades, numerous data analysis tools have become available to researchers working with text. NVivo, Wordstat, LIWC, and T-LAB are some examples of such textual analysis software packages. Although the value of textual analysis programs in social sciences should be self-evident from numerous empirical studies, it is still relatively rare for researchers in education to adopt these tools in their arsenal. This study intends to aid researchers and graduate students in selecting appropriate research procedures and textual analysis tools in their work.

The following section provides an overview of the text mining technique, including terminologies, concepts, typical applications, and some technical details such as the data analysis and data collection techniques. The goal is to prepare readers with some background knowledge to provide a foundation for the research procedures discussed in chapter three.

Defining Terms

The definition of text mining varied slightly among scholars. Feldman (2007) defined text mining as a knowledge-intensive process in which the researcher uses a variety of tools to identify patterns and discover useful information from data sources. Ananiadou and Mcnaught (2006) considered text mining to be an automated technique that is capable of identifying, extracting, managing, integrating, and exploiting knowledge for research and education in an efficient and systematic manner. Feldman and Sanger (2007, p. 1) stated that text mining is a computer-based approach which, "seeks to extract useful information from data sources through the identification and exploration of interesting patterns."

In essence, text mining is the general act of extracting meaningful patterns, themes, trends, or relationships from unstructured text. Some of the commonly used terminologies in text mining include:

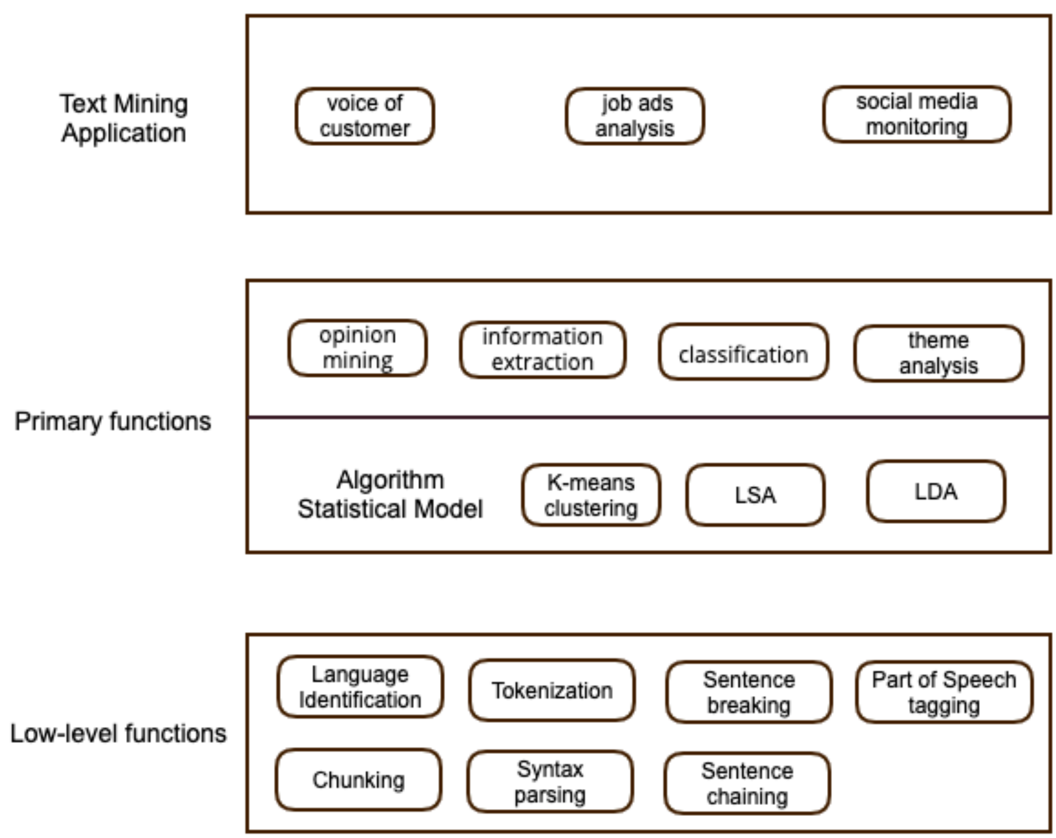
- **Corpus:** The documents that are being selected for analysis. They can be any type of textual material, such as tweets, articles, books, reviews, and comments.
- **N-gram**: An n-gram is a sequence of two or more words that present within the same sentence.
- **Stop word list**: Also known as an exclusion list, the stop word list includes words that need to be excluded in the process of text mining. They are typically common words like “a,” “the,” “and,” etc. These words appear in virtually every sentence but do not have any specific or significant content.
- **Lexical units**: Words and multi-words to be analyzed.
- **Context units**: Portions of text that the corpus can be divided into, such as documents, elementary contexts, and corpus subsets.

- **Occurrences**: The number of times a lexical unit occurs within the context units or a corpus.
- **Co-occurrences**: The number of times two or more lexical units coexist in the same elementary context.

To better organize the terminologies used in text mining, the taxonomy of text mining is presented in a three-layer diagram shown in Figure 1.

Figure 1

Taxonomy of Text Mining



The top layer relates to the research problems. The solution to each problem is often called an application. Some examples are: analyzing online feedback to understand the voice of customers, discovering genres of online discussion, and in this case, analyzing job announcements to identify professional competencies.

The middle layer corresponds to the primary functions of natural language processing, also known as computational linguistics. Natural language processing is the underlying procedure in which computer algorithms can understand the human language, including meaning, tone, emotion, etc. According to Ignatow and Mihalcea (2017), there are four common applications that text mining is capable of: (a) text classification or clustering, (b) opinion mining such as sentiment analysis, (c) information extraction such as named entity recognition, and (4) analyzing topics such as topic modeling or thematic analysis. Typically, each application can be operationalized by a collection of computer algorithms (statistical models). Selecting which algorithm is often a highly technical question.

The bottom layer shows the low-level computational functions that turn raw text into linguistic units and prepare the document for further analysis. These functions form the foundation of natural language processing. Some of the functions are: (a) Language identification, the process of detecting the language of the source document, (b) Tokenization, separating sentences or phrases into smaller units, such as words or lemmas, (c) Sentence breaking, separating each sentence by a punctuation mark, (d) Part of speech tagging, assigning tags to a part of a sentence based on grammar rules, (e) Chunking, further breaking a sentence into small components (phrases), (f) Syntax parsing, the process of deconstructing a sentence, and (g) Sentence chaining, a technique to connect sentences that are related to a central topic.

Primary Functions

Text classification (also referred to as text categorization) is, “the task of assigning texts to one or more predefined categories” (Ignatow & Mihalcea, 2017, p. 285). There is a similar procedure called text clustering. It is important to note that text clustering differs from text classification as it will cluster text into groups without predefined categories. The most commonly used application of text classification is spam email detection. Another example of text classification is called "author profiling," which determines the author's age, gender, or political orientation automatically. This technique can also be applied to detect false reviews and deceptive posts on social media. Jiang and McComas (2014) examined the inclusion of Nature of Science from a group of popular science writings. Their goal was to access the accuracy of text mining and validate whether this technique can be used reliably in science education research. They concluded that text mining techniques improved efficiency and accuracy in determining whether a science writing explicitly includes Nature of Science, and suggested future applications for text mining in analyzing other aspects of science texts.

Opinion mining is defined as the task of identifying private state such as emotions, sentiments, evaluations, beliefs, and speculations in natural language. Usually, there are two types of opinion mining: subjectivity analysis and sentiment analysis. The former identifies if a text contains an opinion by labeling the text as either subjective or objective. The latter classifies the text as either positive, negative, or neutral (Ignatow & Mihalcea, 2017). Amadio and Procaccino (2016) did a study of analyzing text-based online reviews in the hotel industry using text mining tools. They stated that instead of relying on numeric values for hotel rating, text mining could be used to discover information within written comments, revealing the dominant features of each hotel.

Information extraction (IE) is the task of extracting structured information from unstructured data (Ignatow & Mihalcea, 2017). A well-defined subclass of IE is called, "named entity recognition," which extracts predefined information such as people, organization, and location. Another subcategory of IE is Relation Extraction, which typically determines the relationship between two entities.

Analyzing topics is also known as topic modeling. A topic here can be described as a cluster of words that often coexist in the same context. Topic models are statistical models for identifying what combination of topics are discussed within a social group, and how topics discussed change over time (Ignatow & Mihalcea, 2017). Jacobi, Van Atteveldt, and Welbers (2016) conducted a study analyzing nuclear technology coverage in the New York Times from 1945 to 2016. They confirmed that LDA is a useful tool for analyzing news content in large digital news archives with efficiency. Analyzing themes is the most frequently used text mining technique in social science studies. For example, Lin, Hsieh, and Chuang (2009) proposed a genre classification system, called GCS to discover genres of online discussion automatically.

Topic Modeling Algorithm

Topic modeling is the primary function applied in this study, so it is important to discuss briefly how it works. The underlying assumption is that meaning associated with text can be represented using a set of word clusters. Linguists call this a bag of words, which only takes into consideration word occurrences regardless of syntax or context.

Contrary to content analysis, which typically starts with predefined codes or categories, topic modeling usually begins with specifying K, the number of categories the researcher wishes to find. According to Greene, O'Callaghan, and Cunningham (2014), selecting the right K is a highly technical issue. K being too big can result in small redundant topics, while K being too small can lead to overly broad topics.

Once K is set, the computer program identifies the specified number of topics and generates the bag of words within each topic as well as the topic distribution across the text. From an algorithm perspective, topic modeling is an instance of probabilistic modeling, and the most commonly used model is called Latent Dirichlet Allocation (LDA). A natural way to denote a topic is to use a list of words. A word list is capable of describing complicated and abstract topics. By assigning weights to each selected word (term), related words can be grouped together to model a topic and to give a close estimation of what the topic covers. Some words are shared by multiple topics, meaning that these words are related to multiple topics due to their ambiguous nature. Having probabilities of the same word in different topics can address the issue of word ambiguity (Zhai & Massung, 2016).

To give the problem a mathematical description, the input data is a collection of documents, and we can assume there are K topics in the corpus. Here, we will use words as lexical units since words are the most natural units in English. The output data is two types of probability distributions. The first distribution contains a list of topics, and each topic consists of a distribution of all the key terms. The second type of distribution is about topic coverage, where each document is represented by a distribution of all the topics.

Text Mining/Analysis Tools

Textual mining tools come in a variety of forms. Some, such as Diction and Concordance, are tools that can only perform one type of function. Other tools, like QDA miner's Wordstat are all-in-one text mining packages that aim to provide a one-stop solution to their users. These types of tools include:

- **T-LAB**: A comprehensive text analysis and text mining software with a set of linguistic, statistical and visualization tools.
- **Rapidminer**: A data science software platform that supports text mining and text analytics.

- **QDA Miner**: A data analysis software focusing on qualitative research. It features coding, annotating, retrieving, and analyzing textual data.
- **Wordstat**: A quantitative content analysis and text mining software. The main functions include: information extraction, knowledge discovery, automatic tagging, classification of documents, etc.
- **Lexalytics**: A complete software toolkit for analyzing text documents. It is capable of doing sentiment analysis, categorization, entity extraction, theme analysis, intention detection, summarization, etc.
- **SAS Text Miner**: A text mining plug-in for the SAS Enterprise Guide that provides different capabilities for analyzing text.

The third type of tool is a programming language with text mining modules. For example, R is a programming language primarily used for statistical analysis, but it lends itself well to text mining. Python is a programming language that is used by many for text mining and analysis. Researchers typically use either software package or programming language to create a customized text mining solution for their research problems.

Text Mining Studies in Job Analysis

Although text mining as a research methodology has received growing popularity during recent years, the level of adoption varies from discipline to discipline. It seems that text mining, as a research technique, has not been widely accepted as a mainstream methodology in education. Here are some speculations that might explain this disparity. First of all, text mining courses are usually offered by computer science and information science departments rather than educational research programs. Therefore, the lack of familiarity or exposure could be a limiting factor. Secondly, text mining, as a research tool, does have a steep learning curve. Most scholarly publications on this topic are packed with algorithms and mathematical equations, which can be quite overwhelming to novice researchers from a different field. Furthermore, technical and logistical issues can also impact the adoption rate. For example, textual data for mining is not always accessible and may need specific permissions from the content provider. Lastly, it is common that text mining research requires the development of a customized solution, which could undoubtedly intimidate novice researchers without a programming background.

As a result, text mining studies are typically found in computer science or related areas such as informatics, automation, etc. Only a minimal number of text mining studies can be found in the field of education. Yang, Zhang, Du, Bielefield, and Liu (2016) conducted a text mining study analyzing job announcements from the American Library Association between 2009 and 2014 to examine core competencies of librarianship. Maer-Matei et al. (2019) adopted a text mining approach analyzing a huge number of job advertisements to determine skill needs for early career researchers in Europe. To date, most job analysis studies in the field of instructional technology were conducted using content analysis.

Comparison Between the Two Techniques

Content analysis originated from the social sciences. Aureli (2017) stated that content analysis is mainly applied in social sciences such as business, management, and accounting. Text mining, by contrast, originated from the field of computer science, where there is on-going discussion surrounding technical improvements and new algorithms. It is predominantly utilized in hard disciplines, though there are a growing number of applications in the domain of social sciences.

There is some ambiguity in the literature as to whether text mining is one type of content analysis, and more specifically, a computer-aided content analysis. Some scholars consider topic modeling as computer-aided content analysis because manual coding is replaced by the automated process of reliable algorithms (Gaur & Kumar, 2018).

In qualitative content analysis, the knowledge of the researcher determines how text is coded into categories while in text mining, documents are treated as numeric data and are coded and analyzed by the computer algorithm. Content analysis, when operated in a quantitative fashion, appears quite similar to text mining because (a) both have automated processes to replace manual coding, and (b) both share the same theoretical assumption that, “the higher the frequency or occurrence of a specific word or concept in a document, the greater the emphasis is that the author of the document or interviewee places on it”(Aureli, 2017, p. 9).

However, content analysis is a technique that relies heavily on human coders to interpret the meaning and importance (manifested in terms of frequency) of specific topics regardless of whether an automated process is involved. Only content analysis can detect meaning despite word choice or syntax while, “text mining software takes on a simplified view of language that ignores the complexity of semiosis” (Aureli, 2017, p. 9) .

Undoubtedly, the two techniques can lead to different findings. From a methodological perspective, Aureli (2017, p. 23) stated that content analysis and text mining are, “not irreconcilable methods,” which seems to suggest there is a possibility that the two methods can work together.

Due to the limitations of manual coding and the availability of new computer tools, the use of computer programs to facilitate content analysis in a quantitative fashion has become more and more popular. The current study intends to find out whether the findings from using each of the two techniques may crosscheck or complement one another.

Data Collection Techniques

Depending on the nature of the data source, there are different approaches to collecting online data: (a) search-based approaches, (b) API based approaches, (c) web scraping approaches, and (d) direct access approaches. This study mainly concerns web scraping, in which the researcher has several options: (a) use commercial software, (b) purchase a commercial service, and (c) create a customized tool.

Besides technical considerations, the researcher also needs to be very careful with the ethical practice of text data collection. Many websites prohibit web scraping, so it is essential to read the terms of use before conducting data collection.

Chapter Three - Methodology

Research Design

The procedural steps a researcher takes to conduct a research project are called research design (Krippendorff, 2004). This chapter outlines the key components of each procedure as well as the designs used in preparation for the analysis.

Design Rationale

Over the last two decades, job analysis has grown significantly in many academic disciplines, including the field of instructional technology (Kang & Ritzhaupt, 2015). Traditionally, job analysis was conducted using content analysis. This technique has the benefit of full-text examination and is still one of the most rigorous methods in social science. However, it also has the drawbacks of being time-consuming and researcher subjectivity.

In contrast, text mining has the capability of processing large volumes of textual data with computer algorithms, which enables the researcher to discover knowledge and draw inferences from big data. This technique also has the advantage of not being subject to human bias. However, text mining has certain limitations as well. For instance, the results are represented in numbers and can be inexplicit, causing ambiguity in meaning.

The researcher chose to combine text mining and content analysis in the current study because: (a) according to Aureli (2017), content analysis and text mining are reconcilable or compatible methods (both methods have been applied to job announcement studies numerous times and have shown promising results), and (b) when combining two methods in one study, the researcher can obtain the benefits of both and allow them to offset one another's drawbacks.

Content analysis served two purposes in this study: (a) it yielded additional information with rich detail that could not be generated from text mining, and (b) it served as a validation process to evaluate the accuracy of the text mining procedure. In summary, the rationale for this design was to use two different yet compatible methods to enhance the rigor and validity of this study.

Data Collection

Data Source

Job announcements can be found from numerous online sources. Some large job sites offer millions of recent announcements, such as indeed.com, Glassdoor.com, Monster.com, etc. Some have a specific concentration, such as HigherEdJobs and EdTechRecruiting. Table 1 gives examples of job sites and the estimated number of instructional technology positions when accessed on Apr 7, 2019.

Table 1

Job sites and their retrievable data size for instructional technology positions

Job site	Instructional Technology Jobs
AECT jobs	1033 jobs since 2019
ISPI career center	474 recent jobs relate to human performances
ATD job banks	500 recent jobs in corporate and business
Indeed.com	22,095 records were found when searching “instructional technology.” Results are not accurate. Only the first 100 pages (1500 jobs) can be accessed.
Monster.com	11,832 records were found when searching “instructional technology.”
Chronicle.com	3749 records were found when searching “instructional technology,” and 2884 records were found when searching “educational technology”
HigherEdJobs.com	3274 records were found when searching “instructional technology,” and 6025 records were found when searching “educational technology”
EdTechRecruiting	Over 5000 job announcements since 2015

Most of the job sites listed in Table 1 prohibit collecting data without permission in their terms of use. The researcher reached out to two job sites to seek permission for academic use but received no response. Therefore, the researcher only selected one data source that did not require such permission. Since the primary research focus was to examine the efficacy of text mining by comparing results with content analysis, the representativeness of the data source became secondary. Thus, the choice of the source can be justified.

In this study, EdTechRecruiting was used as the data source. Job announcements data was collected from Feb 27, 2015 to Jun 6, 2019, with a total of 5957 records.

Data Collection Tool

The data collection process was conducted with a self-created tool. This tool was developed using Python. In the current job site, a two-level data structure was used. The first level displayed multiple job announcements with basic information (job title, hiring organization, geographic location, post date, and a URL link to the detailed job description). The second level contained the content of job descriptions.

The data collection tool consisted of five subroutines which could be executed in sequential order:

1. Save level-one webpages as HTML files. In this case, each webpage contained 25 job announcements. A total of 270 webpages were retrieved.
2. Parse “title, organization, areas, postdate, URL” from HTML, and save the result to a CSV file.
3. Retrieve all the job description pages according to URL and save them as HTML files.
4. Parse “job description” from HTML, and save the result to the CSV file.
5. Character-level cleaning: remove non-printable characters and extra Returns.

It took over a week to design and develop the tool, then another week for testing and fixing bugs. The testing phase was started from a small scale (first ten records, then 50, then 100). As the tool became more reliable, more data was collected until the entire data set was retrieved.

A good and ethical practice is to minimize the load on host servers. This can be achieved by adding sufficient idling time between each retrieval. Therefore, the data retrieval time (several hours) was prolonged intentionally.

The collected data, a CSV file, contained five fields: (a) job title, (b) hiring organization, (c) areas of responsibility, (d) post date, and (e) job description. The "areas of responsibility" field was a predefined job category that could be used for data screening.

Data Preparation

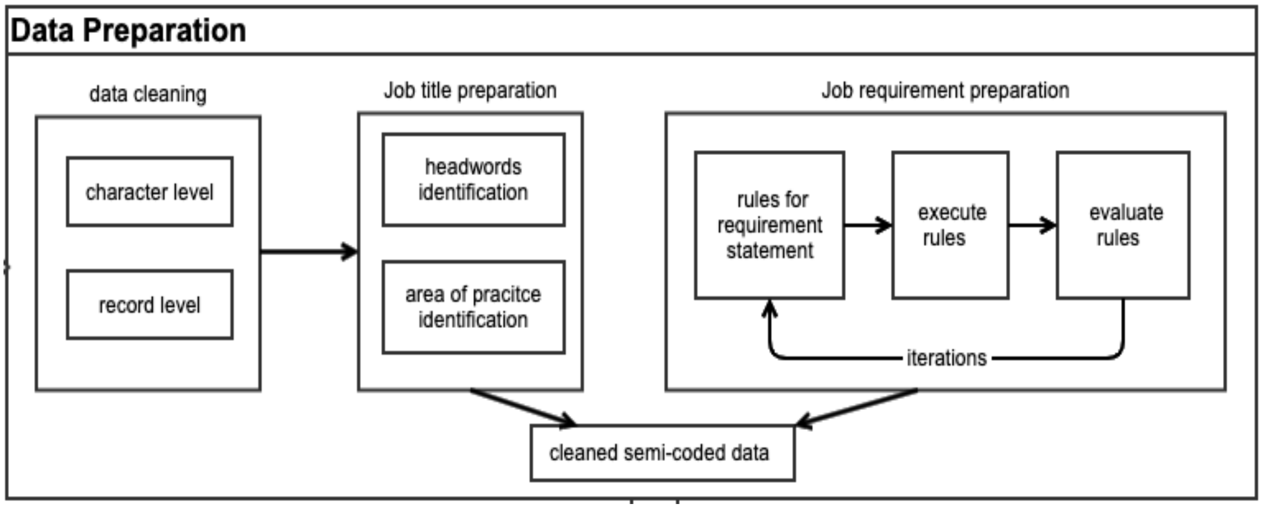
Data preparation is a critical step that affects the outcome of text mining. One big advantage of text mining is the capability of processing a large amount of data within a short period of time. To achieve this goal, every sub step needs to be an automated process. Often times, due to the lack of consistency in the data structure, the data preparation task can be challenging.

The textual data to be analyzed was in the job description field, which contained several types of information: (a) introduction of hiring organization, (b) job responsibility, (c) job requirement, and (d) other information such as how to apply. Job responsibility described the nature of the position and what the applicant was expected to do, while the job requirement stated the expectation from the candidate and what the applicant needed to process. Job requirements included desired qualifications and competencies such as education, work experience, and professional competencies, which could be represented as a combination of knowledge, skill, and ability.

The research design for data preparation is shown in Figure 2.

Figure 2

The research design for data preparation



Following this design, the researcher developed a pre-processing tool using FileMaker Pro 15 to transform raw text to structured data for further analysis. This tool fulfilled two functions: data cleaning and auto-coding. Data cleaning removed duplicate and garbage records from the data set. Auto-coding used pattern recognition with hand-crafted rules to parse and tag data automatically. The primary task was to extract job requirements from the unstructured text.

The purpose of creating and using the pre-processing tool was to reduce the coding time and to enhance the results of text mining by screening the content first. It is expected that this tool can be used for future studies with similar objectives. More details about the data preparation process will be discussed in the following sections.

Data Cleaning Module

A review of the sample revealed enormous diversity and inconsistency within the data set. Since many job announcements were posted by users, inevitably, there were empty, incomplete, and duplicate records. Some records had non-ASCII characters or other unreadable characters. The job description field varied significantly in length. All these issues needed to be addressed before data analysis. Here are the three steps of the data cleaning process:

1. Removing irrelevant records from the data set

Some positions were clearly pure technical jobs by observation. Since the research focus was on instructional technology positions, it was necessary to discard irrelevant records. A job announcement was considered irrelevant if two criteria were met:

- First, the area of responsibility matched one of the following categories: Information Technology, Network/Systems, Registrar, Data/Database, Systems/Operations, and Support/Help desk
- Second, the job title did not have any of the following words: instructional, educational, education, and learning.

2. Removing empty/incomplete records from the data set

Empty and incomplete records needed to be removed because they brought no value to the study. The data cleaning module searched the data set and found 2553 records out of 5956 matching the first criteria. These 2553 positions were then checked using the second criterion. As a result, 2475 job announcements were identified as purely technical and were discarded. The remaining data set contained 3481 records. The pre-processing tool also checked incomplete records and found 112 job announcements had no descriptions; these records were discarded. The remaining dataset contained 3369 records.

3. Removing duplicate records from the data set (de-duplication)

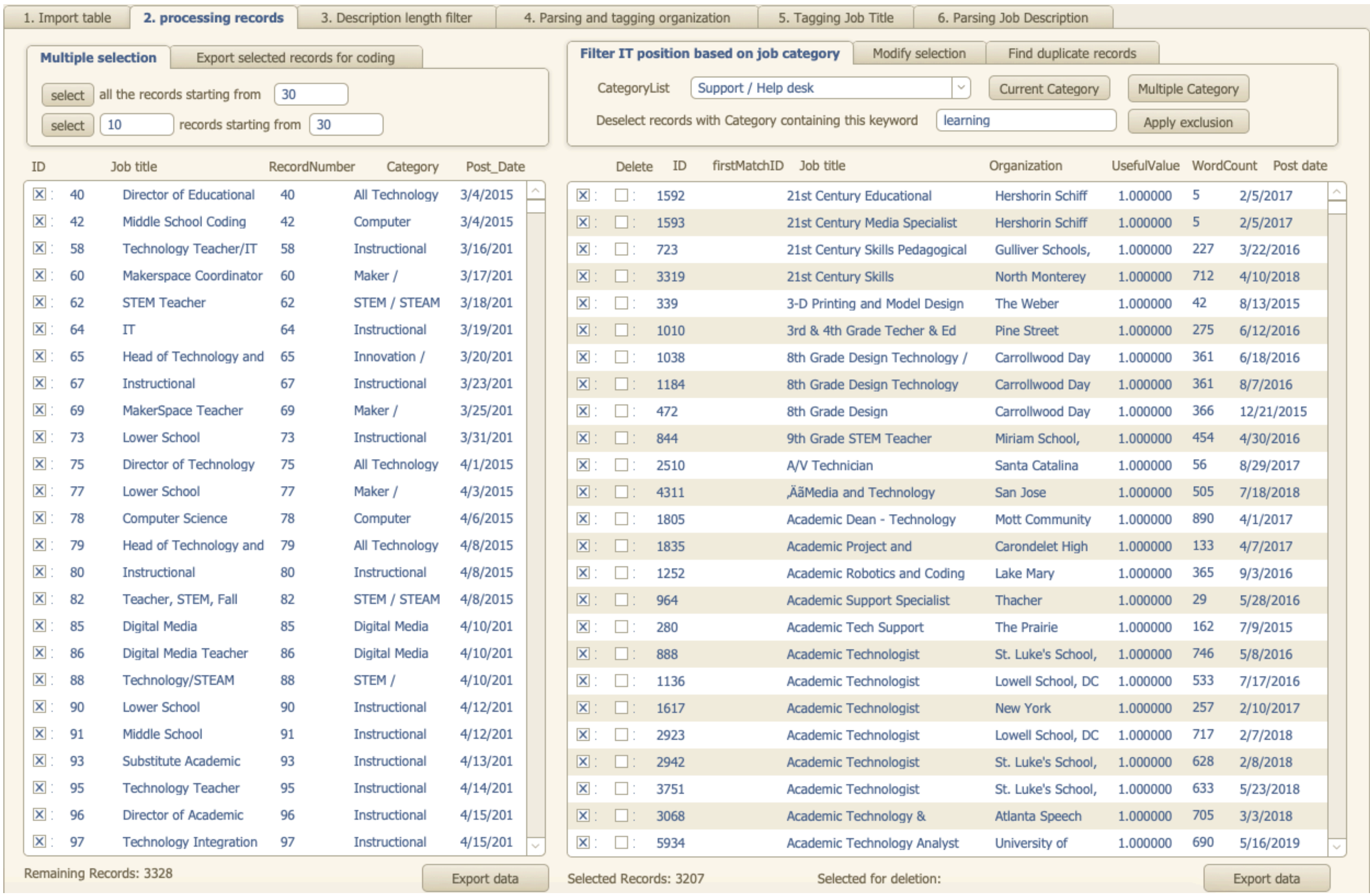
Some job announcements were posted multiple times, causing duplicate records. The final step of data cleaning was to remove duplicate records, also known as "de-duplication"(Carnevale, Jayasundera, & Repnikov, 2014). The data cleaning module used the following criteria to determine duplication: (a) two records have the same job title and organization, and their job description fields have the same length, and (b) two records have the same job title, organization, and post date. The first criteria identified 37 duplicates and the second located three.

After cleaning the data, the data set contained a total of 3328 records.

Figure 3 is a screenshot of the data cleaning module.

Figure 3

Data cleaning module of the pre-processing tool



Job Title Preparation

In English grammar, a headword (or head) is the key word that determines the nature of the phrase which differs from modifiers or determiners (Nordquist, 2015). Each job title is a noun phrase containing a headword. For example, in the title, “digital media specialist,” *specialist* is the headword. Naturally, one way to categorize job titles is to group them by the headwords.

First, all headwords within job titles were identified with the assistance of a text analysis program (T-LAB). T-LAB has a function called “Key Words Selection” (under “Lexical tools” - “Dictionary building/Corpus Vocabulary”). It can analyze all job titles and generate a list of keywords ordered by their occurrence (shown in Figure 4).

Figure 4

Key Words Selection tool in T-LAB

KEY WORDS SELECTION | DICTIONARY BUILDING | CORPUS VOCABULARY

T-LAB	ITEM	OCC
2	TECHNOLOGY	1912
1	TEACHER	687
3	DIRECTOR	664
4	SPECIALIST	500
5	COORDINATOR	354
6	INSTRUCTIONAL	299
7	COMPUTER	250
8	SCHOOL	247
9	INFORMATION	226
10	EDUCATIONAL	211
11	MANAGER	190
12	SERVICE	182
13	SCIENCE	175
14	INNOVATION	162
15	INTEGRATION	158
16	DESIGN	145
17	ASSISTANT	145
18	OFFICER	143
19	MEDIUM	135
20	CHIEF	135
21	DIGITAL	130
22	COMMUNICATION	116
23	ACADEMIC	109
24	LEARNING	99
25	INSTRUCTOR	89
26	MIDDLE	86

By examining the list (183 words), a total of 53 headwords were identified, as shown in Table 2.

Table 2

Headwords identified in job title

Headword	Occurrence	Headword	Occurrence
Teacher	686	Educator	19
Director	668	Administrator	19
Specialist	501	Leader	15
Coordinator	354	Professor	14
Manager	189	Intern	10
Assistant	145	Chair	10
Officer	142	Provost	9
Instructor	90	CIO	8
Associate	86	Trainer	7
Integrator	84	Supervisor	6
Librarian	83	Mentor	6
Engineer	73	Webmaster	5
Coach	70	Chancellor	5
Technologist	58	Architect	5
Support	55	Fellow	5
Designer	54	Editor	4
Registrar	47	Producer	4
Faculty	41	Strategist	3
Analyst	40	Lecturer	2
Developer	38	Tinkerer	2
Technician	37	Advocate	1
Integrationist	29	Steward	1
Dean	26	Operator	1
Consultant	24	Liaison	1
President	23	Advisor	1
Facilitator	22	Collaborator	1
Head	21		

Job Requirements Preparation

After reading the sample text, the researcher found that a job description typically consisted of three components:

- an introduction of the hiring organization (working environment)
- a description of the position: job responsibilities, duties, tasks; that is, what the hire will be doing
- job requirements: competencies, education, and working experience; the personal traits that the ideal candidate needs to possess. Competencies can be further divided into knowledge, skill, ability, and other characteristics.

Since the current task was to examine job-related competencies as the subset of the job requirements, it would be a huge time saver if the job requirement section could be identified and extracted automatically using a computer algorithm.

Therefore, an auto-coding module (classifier) was created to extract job requirements from the data. Here, hand-crafted rules were applied for pattern recognition. The rules were created, tested, and validated through multiple iterations. First, the auto-coding module segmented each job description sentence by sentence. The researcher read the first five sentences and drafted some rules to identify the pattern. These rules were then implemented and tested. The results of computer-coding were evaluated manually. Accuracy measurements were calculated and used as feedback for creating new rules and revising existing rules. This type of iteration continued until satisfactory results were achieved with consistency (five iterations in total). In the field of machine learning, two measures are used to evaluate the accuracy of a classifier: precision rate and recall rate. Precision rate answers the question: Among all identified positives, what proportion was actually correct? It is defined as: Precision = True Positive / (True Positive + False Positive).

Recall rate is the proportion of detected positive in total positive. Mathematically, it is defined as Recall = True Positive / (True Positive + False Negative).

Generally, a precision and recall rate of 0.8 or above is considered satisfactory. In the final iteration, ten untreated records were auto-coded with a precision rate of 0.9375 and a recall rate of 0.9000.

The auto-coding module was operationalized in two steps:

1. Sentence segmentation

The job description field was segmented into sentences based on punctuation marks. These punctuation marks included "; . ? !." However, a period was also found at the end of abbreviations, such as Mr. Mrs. Dr. St. and ex., so these end-of-abbreviation periods had to be replaced in advance to avoid segmentation errors.

2. Rules for requirement statement identification

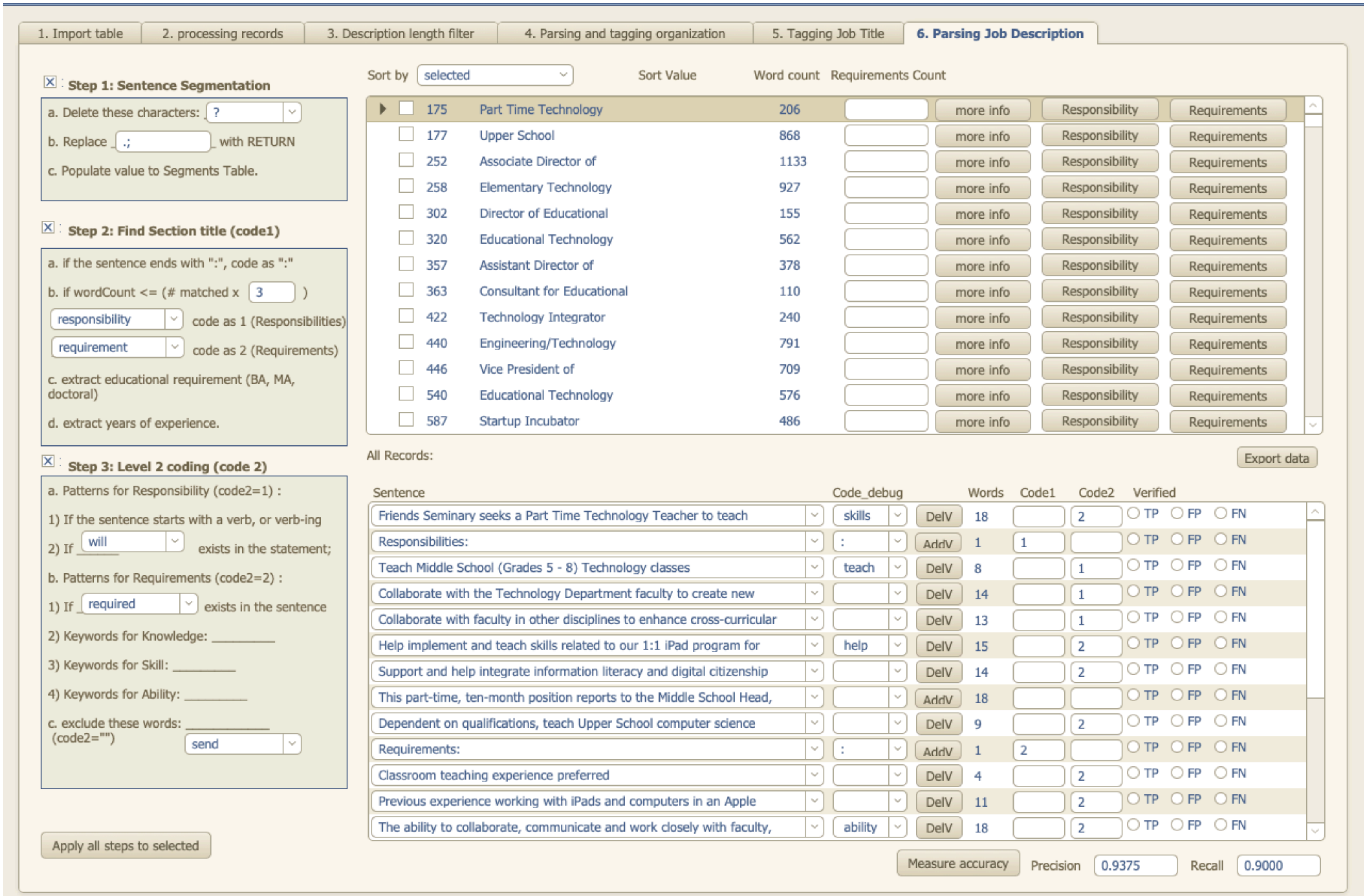
After multiple iterations, the following rules were finalized as an effective classifier to identify a requirement statement:

- A requirement statement must contain one of the following words: required, preferred, preference, expected, should, must, qualification, qualifications, knowledge, skill, skills, ability, abilities, capability, possess, degree, desirable, experience, proficiency, capacity, understanding, ideal, familiarity.
- A requirement statement must have at least four words and does not end with a colon.
- A requirement statement cannot have any of the following words: send, contacted, resume, salary, emailed.

In summary, the pre-processing tool completed three tasks: (a) eliminate irrelevant, incomplete, duplicate data from the sample, (b) automatically code the hiring organization into three categories and code the job title by its headword, and (c) extract the job requirement statements from the job description. All these preparations were crucial to ensure the accuracy of subsequent analyses. Figure 5 is a screenshot of the auto-coding module.

Figure 5

Auto-coding module of the pre-processing tool



The requirement for degree and working experience was considered to be a qualification, not a competency. This type of information was extracted at a later stage using excel formulas when needed.

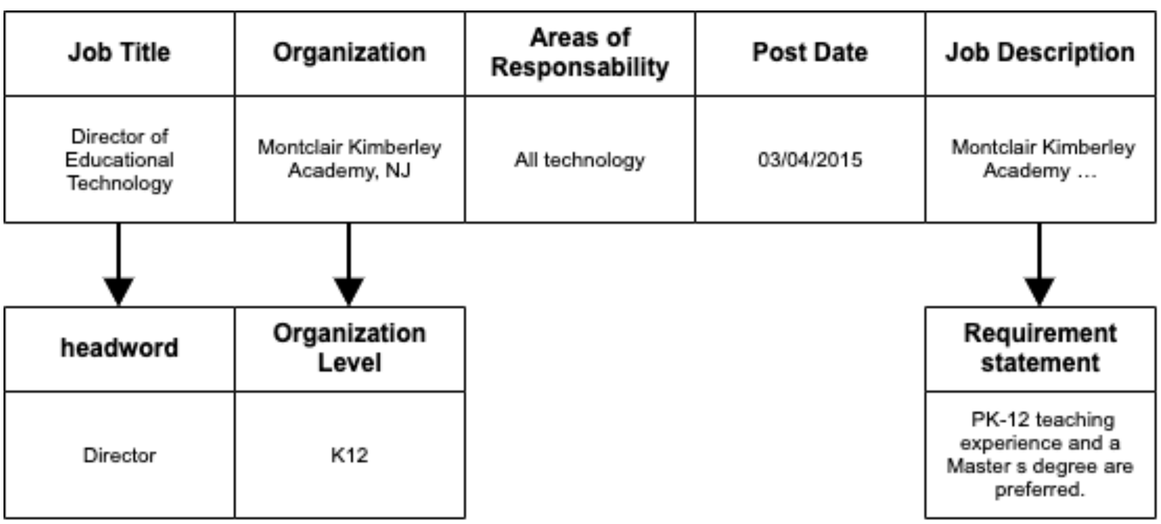
Cleaned and Coded Data

The feasibility of auto-coding relies on the complexity of the analytical constructs. In other words, whether the patterns can be easily captured with explicit rules. In this study, the researcher managed to code job requirements via a rule-based classifier. However, the knowledge, skill, and ability constructs were too complex and abstract to be captured via pattern recognition. Thus, different approaches were used, that is, topic modeling and manual coding.

The data preparation phase successfully created new fields for the data. Figure 6 illustrates the changes made to the data set.

Figure 6

Changes after pre-processing



Content Analysis

Job Title Analysis

The first research question was to identify primary job categories for instructional technologists based on the selected data source. Content analysis was applied to answer the question.

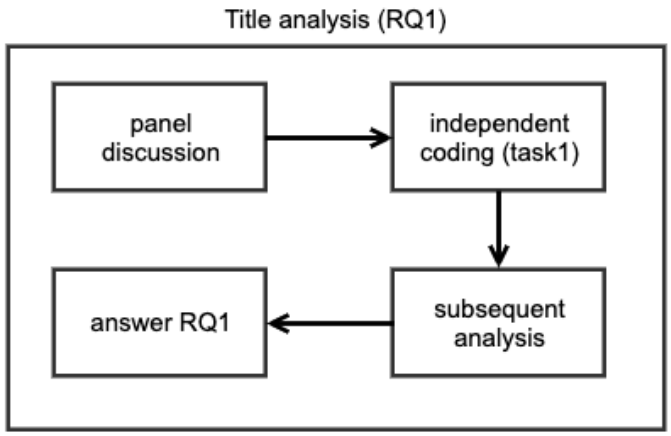
In content analysis, the manual coding process is inevitably subjective. Since the researcher could be too biased to code the content, two colleagues were invited to assist and were trained for the task.

According to Krippendorff (2004), three factors need to be considered when selecting coders: (a) the coders must have sufficient cognitive abilities to map textual units to some type of data language, (b) the coders must have some level of content familiarity or background knowledge to make interpretations based on common understanding, and (c) the level of cognitive abilities and content familiarity should be commonly available within the population of potential coders so that the study can be replicable. The two coders selected were senior Ph.D. students in the field of instructional technology. Both had a sufficient amount of cognitive ability and background knowledge. Their skill level was considered equivalent to their peers.

The job title analysis was conducted in four steps: (1) panel discussion, (2) independent coding, (3) subsequent analysis, and (4) final report, as shown in Figure 7.

Figure 7

The research design for job title analysis



First, the research team had a face-to-face meeting (panel discussion) to introduce the task. The researcher provided basic information and shared training materials—introducing the KSAO framework and how to apply it. Questions and ideas were thoroughly discussed until everyone reached a good understanding of the research task and coding framework.

In coding task one, each coder received a package of two documents: "job title summary report.pdf" and "job title detailed report.pdf." These were generated by the pre-processing tool.

The "job title summary report" was a list of headwords and the occurrences. The "job title detailed report" provided all the instances of each headword. It was anticipated that coders could work independently to develop meaningful job categories.

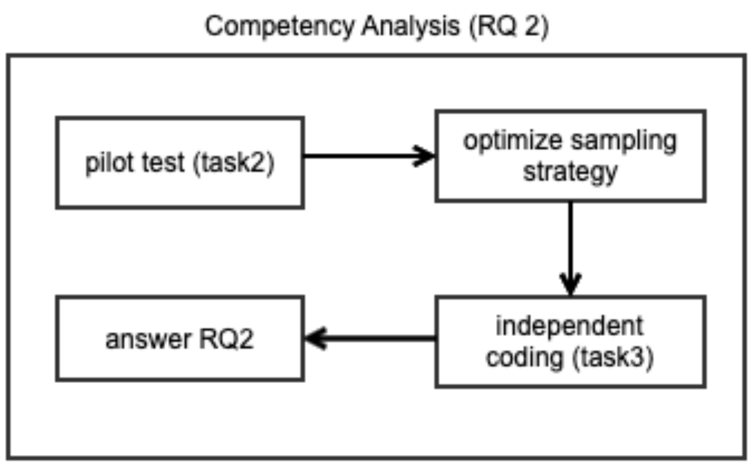
The headwords organization served as a starting point for title analysis. The coders were encouraged to conduct open coding (independent coding). They could either merge several headwords into one big group, or break one headword into multiple subcategories using the detailed report for reference. The researcher compared the findings and completed the subsequent analysis. The results will be discussed in chapter four.

Competency Analysis

Due to the massive volume of content and the abstract nature of the competency construct, conducting a content analysis on job competencies was much more challenging and time-consuming than the title analysis. On average, each job announcement contained 463 words. Due to the limitation of time and resources, it was critical to ensure the research design was sound and plausible. The procedure used to answer the research question two is shown in Figure 8.

Figure 8

The research design for job competency analysis



1. Pilot Test. Manual coding was a labor-intensive and time-consuming process. To estimate manageable sample size to be analyzed, a pilot test (coding task two) was conducted. In this task, each coder received an Excel file (records.xlsx) with ten job announcements (shown in Table 3).

Table 3

Example of coding task two

_ID	segment	code
64	Primary Objective:	
64	To provide campus wide hardware and software support and implement a high quality secondary program in Secondary Technology.	

64	Key Responsibilities:	
64	Plan for and teach High School Technology classes	1
64	Provide hardware and software support to faculty and staff, experience with iPads and interactive white boards desired (Team, SMART, Promethean)	2
64	Implement instructional activities that contribute to a climate where students are actively engaged in meaningful learning experiences	1
64	Provide a positive environment in which students are encouraged to be actively engaged in the learning process	1

Note. _ID: Job announcement document index; Segment: job description separated by each sentence; Code: auto-coding results (1: job responsibility; 2: job requirement) with an accuracy above 0.8.

The coders were asked to determine whether each segment (sentence or phrase) contained any competency statements. If it did, the coders needed to code competency statements according to the KSAO framework.

The pilot test confirmed that: (a) the machine-coded requirement statements were valid sources of job competencies, and (b) analyzing the requirement section alone can yield a sufficient number of competency statements and thus suffice for the purpose of this study.

The pilot test also verified it took about four hours to conduct manual coding on ten job announcements (44 requirement statements).

2. Optimize Sampling Strategy. Based on the pilot test, the following decisions were made for the subsequent analysis:

(1) Sample size. The pilot test identified the average time needed for manual coding. Since 44 statements required four hours of analytical work, a manageable size for the final analysis should be no more than 400 statements (estimated 40 hours of coding).

(2) Sampling strategy. As the number of units to be analyzed was limited, the choice of units became more important to the representativeness of the sample. According to Krippendorff (2004, p. 119), the researcher needs to purposefully shrink the number of

units to a manageable size. In this study, two criteria were used for selecting sampling units: (a) categorical representativeness, and (b) length of the text.

In the previous analysis, fifteen areas of expertise were identified. The researcher and the coders discussed which of these were most relevant to the field of instructional technology and agreed on the following top seven shown in Table 4.

Table 4

Top 7 areas of expertise

Code	Areas of expertise	Common jobs
c3	Instructional Design	Instructional designer
c4	Library/Information Science	Librarian
c5	Multimedia, digital media	Digital media specialist
c6	Technology-infused spaces (maker space, innovation lab, media lab, etc.)	Maker space (teacher/coordinator), lab (coordinator, assistant)
c10	Distance/Online/E- Learning	Digital learning specialist
c12	Assistive Technology	Assistive Technology specialist
c15	Educational Technology/Instructional Technology	Director of educational/instructional technology

Next, computer aids were used to help select representative units. Since the word count of job requirement was calculated in the pre-processing phase, longer documents were in favor. When selecting four records from each of the seven categories, we were able to get a total of 382 recording units, which was reasonably close to the desired number (400).

3. Independent Coding. As a result, 28 job announcements (with 382 requirement statements) were selected as the sample for coding task three. The two coders worked independently on the task, which took them one week (20-30 hours of labor per person). The abridged version of coding results can be found in Appendix A. The findings will be discussed in chapter four.

Text mining (Topic Modeling)

Technical Decisions

Text mining studies require the researcher to close the gap between the research problem and its technical solution. Typically, there are three technical (methodological) decisions to make, as shown in Figure 7.

The first decision is to map the research problem to a primary function that text mining is capable of. In this case, the research question was to identify job competencies latent in job descriptions, which is a typical problem of discovering themes/topic in the text.

The second decision is to select a statistical formula that best fits the current problem. Typically, there is a collection of algorithms or mathematical models for a given analysis. For the current research problem, the algorithms that can perform topic modeling include K-Means clustering and Latent Dirichlet Allocation (LDA).

Both K-means and LDA are unsupervised learning algorithms that are commonly used in topic modeling. Both require the user to select the parameter K. In K-means, K is the number of clusters, while in LDA, K is the number of topics. When both algorithms are executed by assigning K topics to a set of N documents, the difference is: K-means clusters the N documents in K disjoint partitions (i.e., topics in this case), while LDA assigns each document with a topic distribution (Shekhar & Venkatesan, 2018). In other words, each document is characterized by a mixture of topics. For example, document D is comprised of 20% Topic A, 25% topic B, 15% topic C, and so forth. In this study, the research question was to identify job competencies. The goal is to identify themes that match related competencies. The nature of jobs determines that each job will require a mixture of competencies. Hence the LDA model is a better fit for the current research question.

The third decision is to select a text analysis software that can perform all the required computations, which includes low-level text processing, statistical language modeling, and data visualization.

Currently, there are many text analysis or mining tools available. The most robust ones are programming languages such as Python or R. Each has packages and libraries for text mining. These tools are powerful but require strong programming skills. The downside of using a programming language for text mining is that the researcher needs to develop all the low-level computational functions (such as segmentation, lemmatization, vocabulary building), deploy sophisticated statistical models, and take on the risk of possible programming errors. This can be very challenging and labor-intensive. Comparatively, text mining/analysis packages are much easier to use and offer better reliability. For the purpose of convenience and efficiency, the researcher used a text mining package for this study.

Text mining packages are often quite expensive, and can cost thousands of dollars. T-LAB is a relatively affordable choice. It costs €150 ($162) for a one-year student subscription and €330 ($357) for a one-year educational subscription.

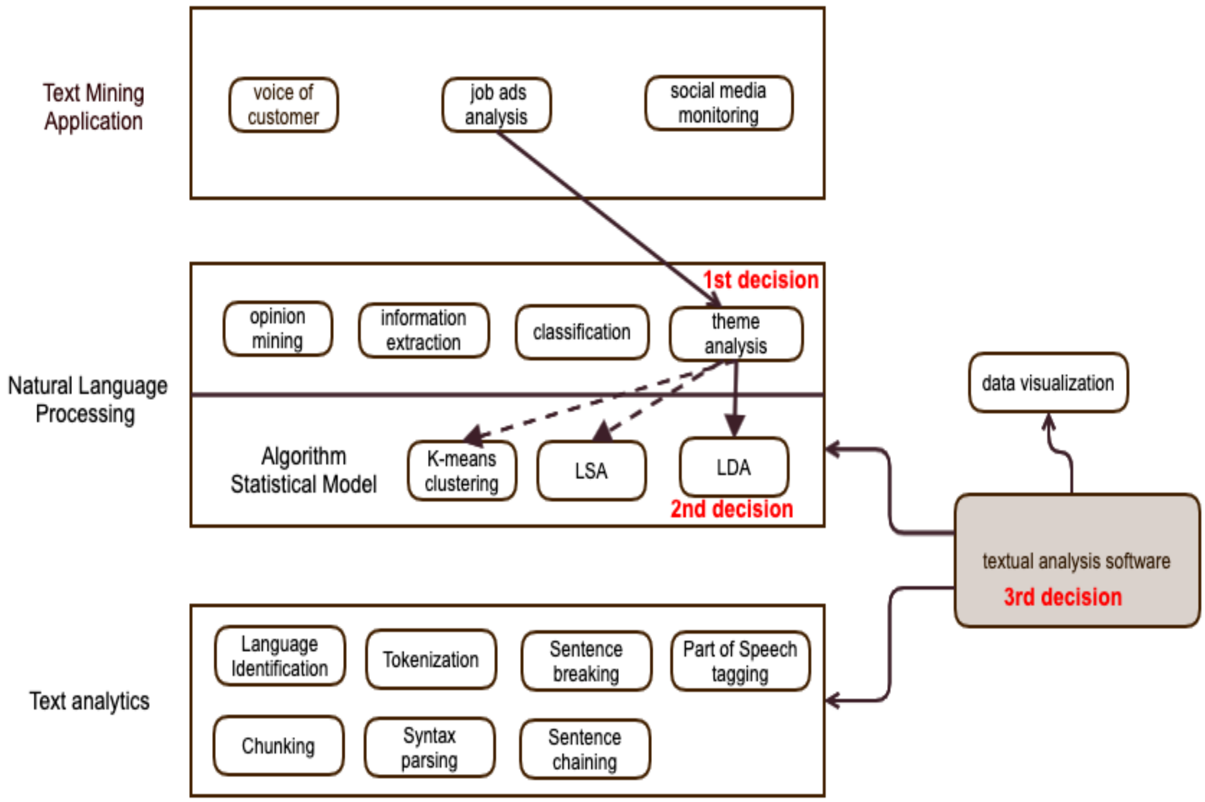
T-LAB includes a set of linguistic, statistical, and visualization tools and focuses on three types of analysis: co-occurrence analysis, thematic analysis, and comparative analysis tools (T-LAB, 2019). These functions were able to satisfy the needs of the current study.

Though still new to many, T-LAB has been used in at least two hundred studies by researchers and professional analysts from over forty countries. Based on these considerations, T-LAB was chosen as the text mining toolkit for the current research.

The three-decision process is illustrated in Figure 9.

Figure 9

Decision process for finding technical solutions

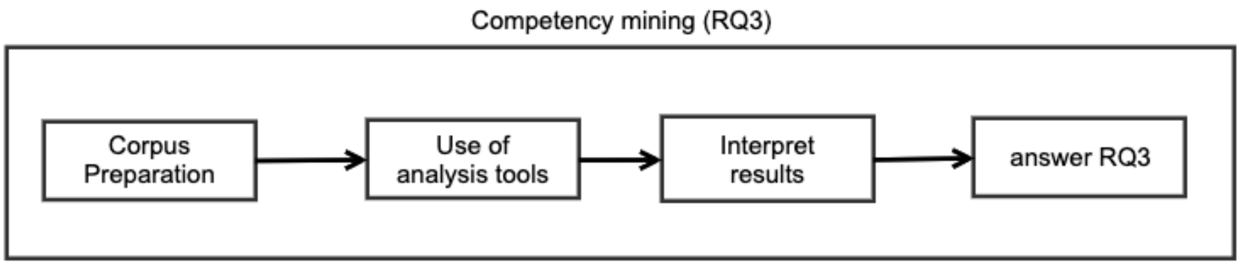


Text Mining Workflow

The workflow of conducting text analysis via T-LAB typically consists of four steps: (a) text gathering, (b) corpus preparation, (c) use of analysis tools, and (d) interpreting results (Figure 10). The text gathering step refers to all the data collection and preparation work that precedes text mining, which has been addressed in the previous sections.

Figure 10

The workflow of text mining



1. Corpus Preparation. The starting point of T-LAB workflow is to load the data. The Corpus Builder tool is a step-by-step wizard to convert and load the data.

According to the T-LAB user manual (T-LAB, 2019, p. 191), there is a set of requirements with which the data needs to comply:

- Each variable can have up to 150 values.
- The IDnumber values, if used, must start from 1, such as 1,2,3, etc.
- Each label, including variable and values, can have up to 15 characters and cannot have spaces.
- Errors will be detected if these requirements are not met.

Here, only three columns from the source data were needed:

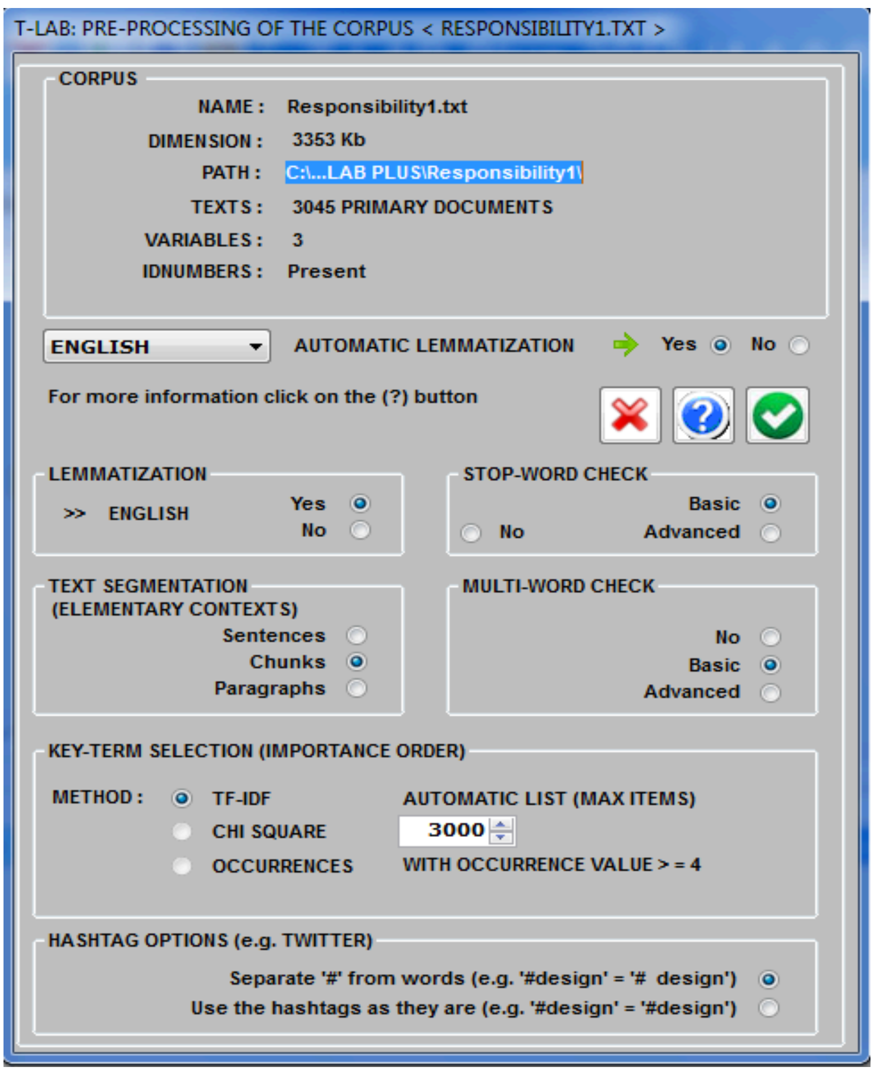
1. ID number
2. Variable (organization)
3. Text (job requirement)

The data set contained 3328 records. Once the Excel file was loaded, the user received an error because some of the recording units were empty. Once 283 empty units were removed, T-LAB was able to proceed to the pre-processing step, as shown in Figure 10. The T-LAB user manual only provides a basic description of the settings; it assumes that users already have the knowledge to understand the terminology and concepts discussed. Hence, having a solid foundation of background knowledge is a prerequisite for the text mining process.

Pre-processing required the user to provide four types of input: (a) lemmatization, (b) stop-word check, (c) text segmentation, and (d) multi-word check. By default, the rule of lemmatization was based on English. A basic stop-word list and a multi-word list were used. Text segmentation was based on Chunks. Chunks are similar to sentences. The only difference is a chunk cannot exceed 500 words.

Figure 11

Parameters used for pre-processing

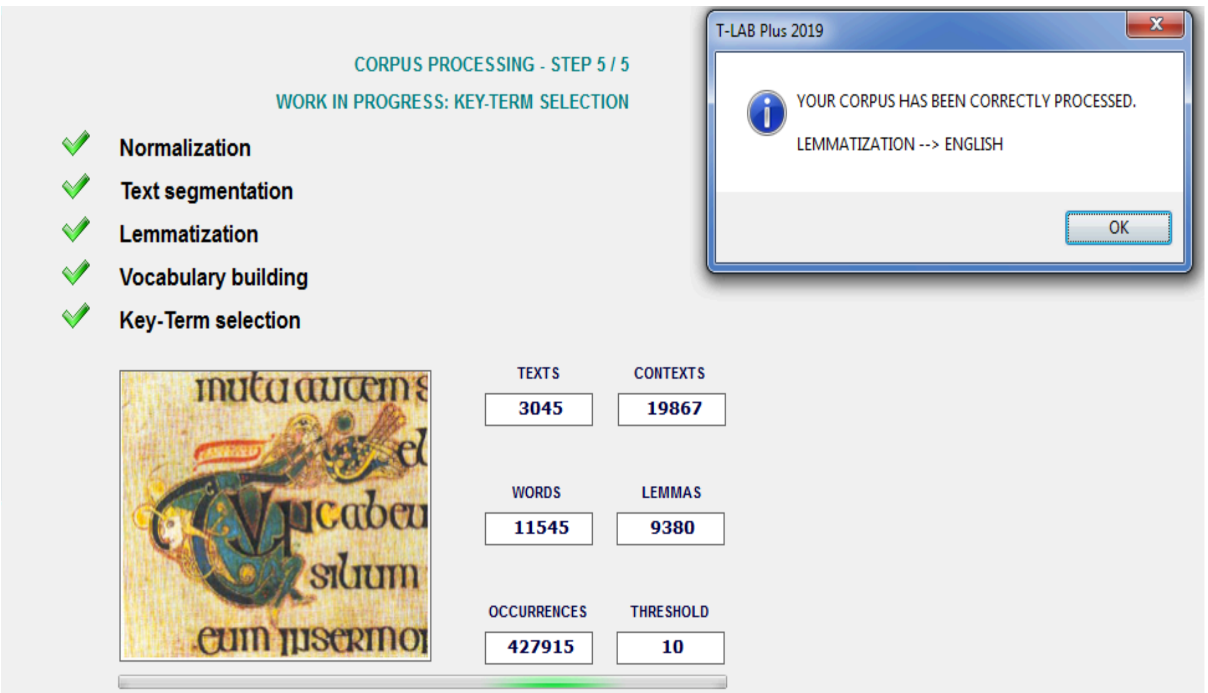


Based on these settings, T-LAB executed a series of sophisticated computations to pre-process the corpus, which were required by the subsequent analyses. One big advantage of using T-LAB was that the user did not need to create any computer algorithms. T-LAB conducted them automatically in the background. These low-level functions included normalization, text segmentation, lemmatization, vocabulary building, and key-term selection.

As a result, T-LAB identified 2036 key-terms from 19866 elementary contexts in 3045 documents (shown in Figure 12)

Figure 12

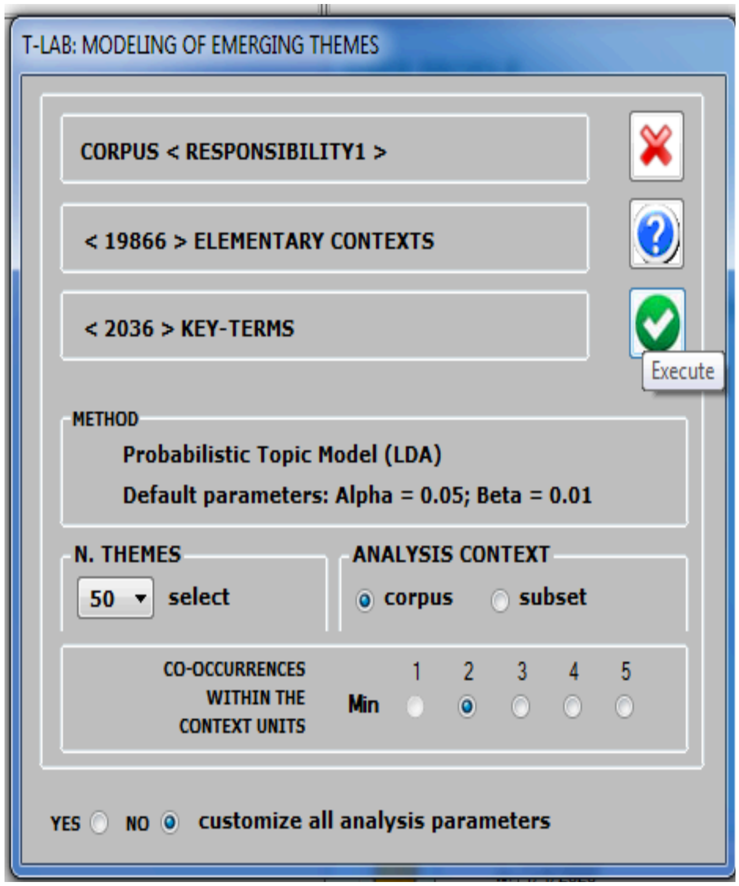
The result after pre-processing



2. Use of Analysis Tools. The analysis tool applied here is called, "modeling of emerging themes," which is based on LDA. Like all unsupervised clustering algorithms, the user needed to set *a priori* - the number of clusters or topics.

Figure 13

Parameters used for modeling of emerging themes



The researcher selected the max number (N=50) that T-LAB could compute. The bigger the number is, the more consistent the coherence patterns are. The themes that are redundant or hard to interpret can be discarded later. Default settings were used for the rest of the parameters (shown in Figure 13).

Then T-LAB executed a series of computations in the backstage which included:

1. Create a document per word matrix. Here documents refer to context units (requirement segments in this case)

2. Analyze data using Latent Dirichlet Allocation model

3. Present themes in terms of probability of the characteristic words (previously identified key-terms). These key-terms can be either "specific" to one theme or "shared" by multiple themes.

Processing time varies depending on the performance of the computer and the volume of the task. In this case, with theme number set to 50, it took about less than 5 minutes to compute the LDA analysis (3045 documents, 19866 elementary contexts, and 2036 key terms). Once the process was completed, the researcher could explore the characteristics of each theme, and decide whether to keep or discard specific themes.

3. Interpret Results. T-LAB provided multiple ways to explore the results (in this case, 50 themes). The researcher found that the "indented tree" format was particularly helpful. The output was an HTML file. It showed all 50 themes as nodes and listed the top 10 most characteristic terms of each node. Based on these ten words, the researcher decided whether to keep or discard the theme based on two criteria: (a) the coherence of the terms – representing a unified construct, and (b) whether the concept reflects one or a group of competencies.

For example, the first theme, with a machine label "Adobe," had the following top ten characteristic words: *Adobe, Captivate, cut, Dreamweaver, Excel, Illustrator, iMovie, InDesign, Photoshop, PowerPoint*. It was obvious that all ten words were related to software programs (meeting the first criterion). This theme can be interpreted as, "the mastery of software skills" (satisfying the second criterion). Therefore, the researcher determined to keep the theme and renamed it, "Software."

The second theme, with a machine label, "Application," had the following terms: *attachment, click, commission, electronically, letter, letters, line, name, packet, resume*. All these words were frequently used in the, "how to apply" section of the job description (satisfying the first criterion). However, this theme was not related to job competencies (failing to meet the second criterion). Thus, this theme was discarded.

The process continued until decisions were made for each of the 50 themes. A complete list of theme interpretations can be found in Appendix B. The findings will be discussed in chapter four.

Chapter Four - Findings

General Description of Job Data

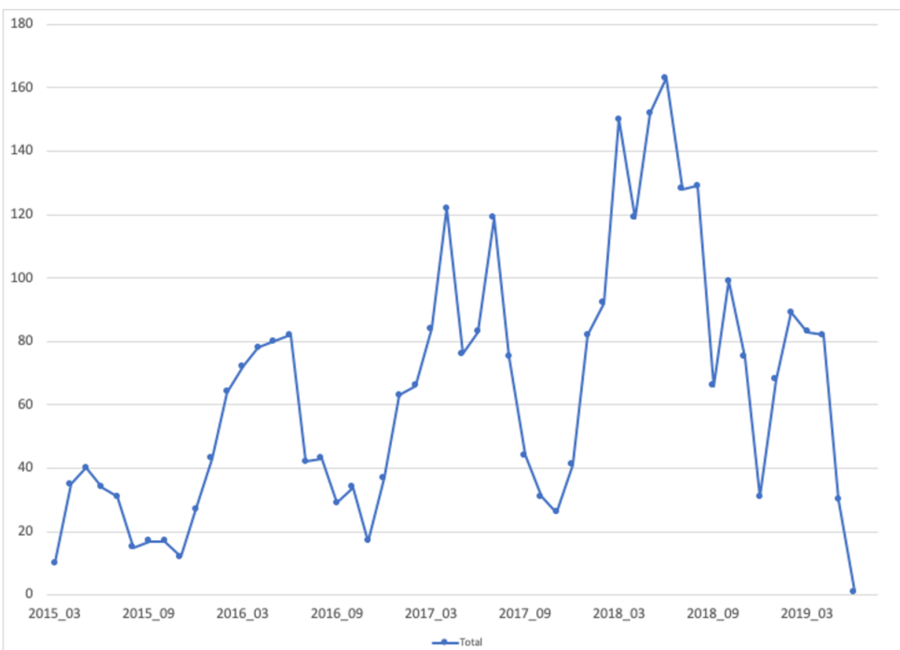
In this chapter, the findings of the job analysis will be discussed. The first section will contextualize the nature of the data set with basic statistical description.

Annual Trend

The job announcements were posted between Feb 27, 2015 and June 6, 2019. The job announcement numbers were plotted by month (Figure 13), making it clear that the peak season occurred between March and August each year, while November or December had the lowest number of job openings.

Figure 14

Annual Trend of Job Announcements



Geographic Distribution

The geographic location data can be extracted from the hiring organization field using formulas in Excel. For example, if cell "H2" is the hiring organization with the value "Montclair

Kimberley Academy, NJ," the formula to extract the text after the comma is "=RIGHT(H2,LEN(H2)-FIND(",",H2))" which, in most cases, returns the abbreviation of a state ("N.J." in this case).

However, there were a few exceptions where the organization data contained multiple commas. In order to resolve the issue, the formula was revised to:

=IF(ISNUMBER(FIND(",", Q2)), RIGHT(Q2,LEN(Q2)-FIND(",",Q2)),Q2)

The data set includes all 50 states in the U.S. as well as some in foreign countries. Geographically, the data showed an uneven distribution where the top six states (California, New York, Texas, Massachusetts, Florida, and Maryland) contributed to roughly 50% of the job openings as shown in Table 5.

Table 5

Number of Job Postings in each State (Top 20)

State	Jobs	Percent
CA	648	19.47%
NY	291	8.74%
TX	219	6.58%
MA	183	5.50%
FL	180	5.41%
MD	138	4.15%
PA	132	3.97%
NJ	119	3.58%
GA	105	3.16%
VA	95	2.85%
WA	86	2.58%
IL	77	2.31%
NC	68	2.04%
CO	67	2.01%
HI	65	1.95%
OH	56	1.68%
CT	49	1.47%
IN	47	1.41%
DC	41	1.23%
MO	35	1.05%

Contextual Distribution

Upon close examination of the context, it seems that all positions were in the education setting, and could be classified into three contextual categories: K12 education, higher education, and others.

The following codebook (Table 6) was used to code contextual categories based on matching keywords:

Table 6

Codebook for categorization hiring organization

Context	Keywords	Percent
K12 Education	School, Elementary, Preparatory, ISD, School District, Academy	67.6%
Higher Education	University, Institute, College	27.5%
Others	None of above	4.9%

This procedure was completed using the auto-coding module in the pre-processing tool.

The data showed that all positions were in the field of education: 67.6% of jobs were in K12 education, 27.5% were in Higher Education, and the remaining 4.9% could not be classified using the current codebook.

1. How might job categories be identified when using multiple approaches to aggregating job titles?

In order to answer the first research question, two coding strategies were applied to analyze job titles, resulting in two categorizing systems: one based on headwords and the other based on areas of expertise. By applying multiple approaches, we might have a better understanding of the multi-faceted nature of job categories.

Job Categories Based on Headwords

The researcher first conducted job title analysis using T-LAB and identified a total of 53 headwords. The occurrence of each was calculated using the pre-processing tool, and the results were shown in Table 2.

The usage of the headwords varied greatly. Some were heavily used, such as “teacher,” “director,” “specialist,” “coordinator,” and “manager,” which could indicate a high demand for such positions. Others, on the contrary, had very low occurrences: it could be the case that such positions were rare, like “steward,” or “liaison.” It could also be the result of novice use of buzzwords such as “tinker,” or simply a rare word choice such as, “integrationist.”

Four headwords exceeded the 10% threshold in K12 education. Top five job titles for each group were identified. Table 7 presents the most common career opportunities for instructional technologists working in K12 education.

Table 7

Common job categories in K12 education

Headword	Frequency	Job title
Teacher	28.2%	Technology teacher Science teacher Computer teacher Robotic teacher Design teacher
Director	19.4%	Director of technology Director of instructional technology Director of educational technology Director of education technology Director of innovation

Specialist	13.9%	Technology integrate specialist Technology specialist Media specialist Support specialist Innovation specialist
Coordinator	10.8%	Education/Educational Technology coordinator Academic Technology coordinator Program coordinator Innovation coordinator Technology integration coordinator

In higher education, four headwords exceeded the 10% threshold. Associated job titles were identified. Table 8 lists the most common career opportunities for instructional technologists working in higher education.

Table 8

Common job categories in Higher education

Headword	Frequency	Job title
Director	20.8%	Director of information technology Director of technology Director of education technology Director of learning technology
Specialist	14.4%	Educational technology specialist Instructional design & technology specialist Instructional design specialist Academic technology specialist Training specialist
Officer	10.9%	Chief Information officer Chief Technology officer Information technology officer Information Security officer
Manager	10.1%	Project manager Service manager Technology manager Security manager Program manager

Naturally, one way to further aggregate job titles was to group similar headwords into broad categories (inductive approach). Table 9 presented one example of such a strategy, which resulted in six major categories based on mapping headwords to each group.

Table 9

One approach to grouping headwords

Leadership	Teaching	Training	Supportive	Creative	Technical
Director	Teacher	Trainer	Coordinator	Designer	Technician
Head	Educator	Coach	Integrator	Editor	Assistant
Supervisor	Instructor	Mentor	Facilitator	Producer	Support
Manager	Faculty		Integrationist	Developer	Registrar
Leader			Specialist	Webmaster	Engineer
Dean			Technologist		Analyst
President			Advocate		*Administrator*
Associate			Support		*Officer*
Administrator			Librarian		
Officer			Consultant		

Job Categories Based on Areas of Expertise

However, this categorizing strategy could be ineffective if headwords did not accurately represent essential job functions. Sometimes, a headword was categorized differently in different contexts. For example, "administrator" and "officer" could both be found in leadership positions and technical positions depending on the context. The issue of word ambiguity could make such a strategy problematic.

In addition, the process of grouping headwords was subjective, which could make the results questionable. As such, coding task 1 was created to address the issue. This task was designed to see how human coders could categorize job titles and potentially find new strategies to categorize job types.

As a result, one coder completed task 1 with a creative approach. This coder identified fifteen areas of expertise from the data and presented a crosstab visualization of the job categories with areas of expertise, as shown in Appendix C. Table 10 is the abridged version of Appendix C, which presents the relations between the most common headwords and areas of expertise. As shown in the table, some headwords, such as “manager,” “director,” and “specialist” are associated with a wide range of expertise. Therefore, it is hard to predict the job functions based on the job title.

Table 10

Relation between the headwords and areas of expertise

Area of expertise	officer	manager	teacher	director	specialist	coordinator
Web Development, Management, and Service		x		x	x	x
LMS		x				
Instructional Design/Educational Technology/Instructional Technology	x	x	x	x	x	x
Library/Information Science	x	x		x	x	x
Multimedia		x	x	x	x	x
Technology-infused spaces (maker space, innovation lab, media lab, etc.)		x	x	x	x	x
Information Technology/Computer Science/programming	x	x	x	x	x	x
Marketing and Communications	x	x	x	x	x	x
STEM/STEAM			x	x	x	x
Distance/Online/E- Learning		x		x	x	x
Research		x	x	x		x
Assistive Technology				x	x	x
Social media					x	x
Project Management		x		x	x	
Administration		x		x	x	
Educational Information/Records Management and Service		x		x	x	

In summary, to answer research question one, two coding strategies were applied to conduct content analysis on job titles, resulting in two categorization systems: one based on headwords and the other based on areas of expertise. Job types based on headwords categorization were reported. The relations between headwords and areas of expertise were also presented.

2. What Instructional Technology Competencies might be Identified via Human-based Content Analysis?

In this study, instructional technology competencies were analyzed using the KSAO framework. Therefore, the findings were broken down into four sections: (a) domain of knowledge, (b) domain of skill, (c) domain of ability, and (c) other competencies.

Domain of Knowledge

The two coders identified 27 and 25 knowledge statements respectively. The researcher compared the codes and found that 12 of them overlapped - that is, they used different words to express the same meaning. After merging and re-organizing the codes, the researcher came up with 20 competency statements as shown in Table 11. Sub-categories identified by the coders were also reported.

Table 11

Competency in knowledge domain

Codes	Statement
K1	Knowledge of instructional technology
K2	Knowledge of learning theories
K3	Knowledge of instructional design
K4	Knowledge of children and young adult literature
K5	Knowledge of library and information science
K6	Knowledge of distance education
K7	Knowledge of trends in academic research support
k8	Knowledge of quality assurance framework such as Quality Matters

K9	Knowledge of operating systems
K10	Knowledge of the integration of culture and curriculum
K11	Knowledge of mass media communication, e.g., social media
K12	Knowledge of culture-based curriculum
K13	Knowledge of pedagogical practices, trends, and assessment
K14	Knowledge of technology integration
K15	Knowledge of 21st century skills
K16	Knowledge of LMS
K17	Knowledge of pertinent laws and policies such as common civil rights laws
K18	Knowledge of multimedia production software
K19	Knowledge of information technology
K20	Knowledge of trends in educational technologies, teaching, and learning.

Sub-categories of K1 - Instructional Technology	
K1.1	Google apps for education
K1.2	Web technologies
K1.3	Interactive tools
K1.4	Assistive technology
K1.5	AR/VR/MR technologies
K1.6	Maker spaces

Sub-categories of K19 - Information Technology	
k19.1	Knowledge of 3D modeling and fabrication technologies
k19.2	Knowledge of Microsoft Office Suite
k19.3	Knowledge of data management and analysis software
k19.4	Knowledge of mobile and web development
k19.5	Programming

Domain of Skill

When codebooks were compared, both coders identified 18 skill statements, with seven of them overlapping. The two codebooks were compared and merged, resulting in the 23 skills identified and shown in Table 12. Sub-categories identified by the coders were also reported.

Table 12

Competency in skill domain

Codes	Statement
S1	Written and oral communication skills
S2	Supervision skills
S3	Computer and technical skills
S4	Operation skills
S5	Interpersonal skills
S6	Problem-solving skills
S7	Observational skills
S8	Organization skills
S9	Presentation skills
S10	Training skills
S11	Multitasking skills
S12	Leadership skills
S13	Team building skills
S14	Collaboration skills
S15	Customer service skills
S16	Project Management Skills
S17	Time management skills
S18	Data analysis skills
S19	Critical thinking skills
S20	Facilitation skills
S21	Curriculum design skills
S22	Visual Design skills
S23	Maker skills

Sub-categories of S3 - Computer and technical skills

S3.1	Microsoft office skills
S3.2	Digital tool implementation skills

Sub-categories of S3 - Organization skills

S8.1	Event planning skills

Domain of Ability

Two coders identified 17 and 40 abilities, respectively. Coder B (the one with 40 codes) had attitudinal and motivational statements mixed with ability statements. The researcher compared and merged similar codes. The final codebook for ability domain includes 22 ability statements as shown in Table 13. Associated sub-categories were also reported.

Table 13

Competency in ability domain

Codes	Statement
A1	Support instructions using assistive technologies
A2	Teaching
A3	Work with varied stakeholders
A4	Work with colleagues, collaborate in a team
A5	Develop and manage lab space
A6	Design, develop, and implement trainings and learning solutions
A7	Carry out administrative responsibilities
A8	Foster design mentality
A9	Design and develop curricula
A10	Provide instructional and technological consultation and support
A11	Deal with sensitive and confidential issues in a professional manner
A12	Use and promote educational technologies
A13	Identify and apply appropriate methods
A14	Keep current of distance education and learning technology
A15	Manage high-risk initiatives
A16	Design, develop, and assess standards-based programs
A17	Evaluate technology integrations and learning programs
A18	Manage competitiveness and conflict in work situations
A19	Use and manage learning management system (LMS)
A20	Work with individuals of special needs
A22	Perform tasks that require reasonable physical abilities

Sub-categories of A2 - Teaching	
a2.1	Apply instructional strategies to accommodate diverse learning styles and age groups
a2.2	Use class management techniques

Sub-categories of A3 - Working with varied stakeholder

A3.1	Collaborate with individuals of various age group and diverse academic disciplines

Sub-categories of A7 - Carry administrative responsibility	
A7.1	Organize and prioritize task
A7.2	Schedule meetings and activities
A7.3	Manage schedule and meet deadlines
A7.4	Train, supervise, and evaluate staff
A7.5	Develop and implement safety, maintenance, and tracking systems

Sub-categories of A10 - Provide instructional and technological consultation and support	
A10.1	Work with audience of varied technology capabilities

Other competencies

Thirteen statements of other competencies were identified and presented in Table 14. These competencies were related to attitude, motivation, belief, or religion.

Table 14

Competency as other characteristics

Code	Statement
O1	Self-directed and highly organized
O2	Adhere to organization's mission
O3	Foster innovation among communities
O4	Energetic
O5	Willingness to continue professional development
O6	Passion for work
O7	Proactive and self-motivated
O8	Empathy, resilience, and humor
O9	Strong work ethic
O10	Detail-oriented
O11	Living examples of the institution's value
O12	Promote love and respect
O13	Relevant religious belief

3. What Instructional Technology Competencies might be Identified from Computer-based Text Mining?

To answer this question, a text mining strategy (topic modelling) was applied. The outcome of this approach was called topics or themes. 50 themes were first identified by computer using LDA algorithm, then reduced to 20 by the researcher. Semantic coherence of each of the 20 themes was assessed, and examples of typical competency statements were presented.

1. Identified themes

Based on the researcher's input, T-LAB computed the modeling of emerging themes (LDA topic modeling) and identified 50 themes. The researcher evaluated each theme, kept 20 themes (shown in Table 15) corresponding to required competencies, and discarded the rest. The themes were labeled based on the researcher's interpretation, and categorized using Bloom's domains of learning (Bloom, 1956).

Table 15

Identified themes

No.	Theme	Domain
1	Cultural Values	Affective
2	Personality	Affective
3	Religious Belief	Affective
4	Soft Skill	Affective
5	Work Style	Affective
6	Physical Mobility	Psychomotor
7	Data Management and Analysis	Cognitive
8	Design and Production	Cognitive
9	Hardware	Cognitive
10	Instructional Theory	Cognitive
11	Library and Information Science	Cognitive
12	Learning Management System	Cognitive
13	Marketing	Cognitive

14	OS and Networking	Cognitive
15	Problem Solving	Cognitive
16	Robotic and Coding	Cognitive
17	Privacy and Security	Cognitive
18	Software	Cognitive
19	Teamwork and Relationship	Cognitive
20	Design Thinking	Cognitive

2. Assess the semantic coherence of each theme (quality measure)

The quality of the themes can be measured with the semantic coherence index using T-LAB's Quality Indices (Table 16). The quality index represents the semantic quality of each cluster, which is computed by pairwise similarities between the top ten words of each theme. The top 10 words are those with the highest probability values of each theme, and the average similarity is computed using the cosine index of each word pair at elementary context level. As shown in Table 15, "Soft skill" has the highest average quality index indicating a high semantic quality based on the following ten words: time-management, prioritization, attention, multi-tasking, proofreading, decision making, detail, oral, organizational, and verbal.

Table 16

Quality Indices of the identified themes

THEME	AVERAGE	MIN	MAX
SOFT SKILL	.027	.003	.067
PERSONALITY	.015	.001	.075
INSTRUCTIONAL THEORY	.013	.003	.03
ROBOTIC AND CODING	.013	.004	.04
THINKING AND MINDSET	.012	.002	.027
WORK STYLE	.012	.002	.03
PROBLEM SOLVING	.01	.002	.045
HARDWARE	.01	.002	.032
OS NETWORKING	.009	.002	.028

TEAMWORK RELATIONSHIP	.009	.001	.037
DESIGN AND PRODUCTION	.009	.001	.037
MARKETING	.009	.001	.02
LMS	.008	.001	.021
PHYSICAL MOBILITY	.008	.001	.036
RELIGIOUS BELIEF	.008	.002	.022
LIBRARY AND INFORMATION SCIENCE	.008	.002	.025
SOFTWARE	.008	.002	.021
CULTURAL VALUES	.007	.001	.021
DATA MANAGEMENT AND ANALYSIS	.007	.001	.017
SECURITY PRIVACY	.006	.001	.026

3. Typical competency statement

Since the computer cannot interpret meanings in context or summarize ideas like human readers, the interpretation of text mining results must come from the researcher. T-LAB can generate multiple outputs to facilitate the process of theme interpretation. The indented tree view is a simplified report which lists the top ten most characteristic terms for each theme. A more detailed report lists the top 20 elementary contexts for each theme sorted by weighted descending order. Table 17 is an abridged version of such a report. The second column (Theme) are themes chosen and relabeled by the researcher. The third column (statement example) are the computer selected elementary contexts with the highest weight.

Table 17

Meaningful context of each theme

No	Theme	Statement example
1	Cultural values	Understanding of, sensitivity to, and respect for the diverse academic, socio-economic, ethnic, religious, and cultural backgrounds, disability, and sexual orientation of community college students, faculty and staff.

2	Personality	The ideal candidate is a metrics driven, collaborative and creative problem-solver share a passion for working with young people
3	Religious belief	Possess a strong Christian commitment and to contribute to the overall life of the school.
4	Soft skill	Strong interpersonal skills Strong verbal and written communications skills Excellent communication and presentation skills both oral and written Excellent organizational skills
5	Work style	Demonstrated ability to work as part of a team, ability to work flexibly and creatively in a changing and fast paced environment with a diverse population
6	Physical mobility	The ability to lift 30 pounds The ability to stand for periods of time The ability to stoop, bend and reach repetitively The ability to walk multiple times across campus
7	Data management and analysis	Knowledge of principles and methods of systems and business process analysis and project management; Knowledge of principles, practices and techniques of information systems management, including applications design, hardware and software options for administrative, business and academic functions and the cost-benefit of systems alternatives;
8	Design and production	Experience using 2D and 3D design software and technology to support arts foundations and exercise digital fabrication techniques
9	Hardware	ability to create and implement network design and installation of hardware and software; ability to define and recommend micro-computer hardware and software; ability to detect and resolve computer equipment and software breakdowns and problems
10	Instructional theory	Knowledge and understanding of effective strategies for instructing diverse learners Knowledge and understanding of effective technology integratio n strategies and research-based best practices for curriculum integration
11	Library and information science	knowledge of current technologies, platforms, and products that support library information technology Knowledge of current issues and trends in collection strategies, collection management, copyright, and preservation facing libraries

12	LMS	Deliver training pertaining to instructional technology systems, managing an online education environment, or administering course or learning management systems. Experience using project management tools, learning management systems, and eLearning development tools
13	Marketing	social media for digital storytelling Strong knowledge of best practices in online and digital communications Preferred experience in digital marketing, analytics, and social media
14	OS networking	Experience working with technology in an academic environment Familiarity and experience with Active Directory, Routed IP Networks, Web Servers, Linux and Windows Servers, DHCP and DNS
15	Problem solving	Ability to determine people who are critical to accomplishing results, and brings them together to problem solve or share workload; identify problems and develop logical conclusions and effective solutions;
16	Robotic and coding	Familiarity with common robotics programs Knowledge of the Java programming language or other object-oriented language
17	Security privacy	techniques of management controls and information security protections In-depth knowledge of security policies and procedures
18	Software	Strong working knowledge of the Adobe Creative Suite be proficient with Microsoft Office Suite, Adobe Creative Suite
19	Teamwork relationship	Demonstrated ability to communicate with users/co-workers concerning complex issues Ability to build and maintain positive relationships with faculty and staff
20	Design Thinking	Implementing hands-on, child-centered learning Experience in Design thinking

In summary, among the 50 machine-generated themes, the researcher selected 20 themes that were most relevant to instructional technology competencies. These themes were grouped using Bloom's domains of learning, and typical competency statements of each theme were also reported.

4. How do the Results of Text Mining Differ from Human-based Content Analysis?

Based on the findings from research questions two and three, similarities and differences will be discussed. The focus is to compare the differences in terms of granularity, comprehensibility, and objectivity.

Content analysis revealed 20 competencies in the knowledge domain, 22 in the ability domain, 23 in the skill domain, and 13 in others, while text mining identified 20 themes/topics for required competencies.

Each mined theme can find a match with at least one of the codes from content analysis, as shown in Table 18. Therefore, the two methods seem to produce outcomes that can be used to cross-check one another. However, the findings also differ in levels of generalization. In most cases, text mining yields a high-level generation in which each theme comprises a cluster of competencies. A potential risk of the theme being too broad is oversimplification or missing important details. For instance, the theme, "soft skill" corresponds to 13 competencies identified by human coders; the theme of "instructional technology" matches nine competencies comprising all theories and practices related to teaching and learning. Obviously, content analysis outperforms text mining in terms of richness and thoroughness.

On the other hand, a small number of content analysis results cannot find a match with text mining themes, such as statements K14, A1, A10, A12 A17, which are all about technology integration. It is likely that this topic was among the 30 themes being discarded. The theme "Support" (in Appendix B) also has, "Tech" listed as a keyword, but the rest of the words cannot be easily interpreted as a coherent theme.

To summarize, text mining results differ from content analysis in three aspects: First, they differ in the level of granularity. In this case, themes identified by text mining have a high level of granularity, that is, each theme represents an aggregation of similar competencies. The product of human coding, though time consuming, can capture job competencies in a more thorough and accurate fashion. Second, they differ in the level of comprehensibility. Each text mining theme in its original form, is a probability distribution of characteristic terms. Some themes are not comprehensible or meaningful to human readers. Others need to be interpreted and synthesized by the researcher so that readers can understand. In short, the competencies of content analysis are in a natural language that can be directly understood by readers, while text mining results require some human interpretation to clearly articulate meaning. Finally, they differ in the level of objectivity. The researcher believes that one big advantage of text mining is its objectivity. Text mining results (themes of competencies) are products of statistical computation defined by the mathematical model and the selected parameters. They are considered more objective because of minimal human manipulation. For example, problem solving skill is considered to be a subordinate of "Soft Skills." However, the computer algorithm identified it as a separate theme. From a statistical perspective, this means that problem solving skill is significant enough to be considered an independent category. This unbiased finding is helpful for readers to understand the significance of theme.

Table 18

Results compared between text mining and content analysis

No.	Text mining theme	Content analysis code
1	Cultural Values	O2, O3, O11, O12
2	Personality	O4, O8
3	Religious Belief	O13
4	Soft Skill	A15, S1, S2, S4, S5, S7, S8, S9, S10, S11, S17, S19, S20

5	Work Style	O1, O5, O6, O7, O9, O10
6	Physical Mobility	A22
7	Data Management And Analysis	S18
8	Design And Production	K18, A5, A6, S21, S22, S23
9	Hardware	K1, K19
10	Instructional Theory	K2, K3, K4, K6, K10, K13, K20, A14, A22
11	Library And Information Science	K5
12	Learning Management System	K16, A19
13	Marketing	K11
14	OS and Networking	K9, K19
15	Problem Solving	S6
16	Robotic and Coding	K19
17	Privacy and Security	A11
18	Software	K1, K19, S3
19	Teamwork and Relationship	A3, A4, A18, A20, S12, S13, S14, S15
20	Design Thinking	A8

Chapter Five - Discussion

In chapter four, both content analysis and text mining were applied to the same data set to extract job competencies. Similarities and differences of the results were compared, and pros and cons of each methodology were discussed. In this chapter, the challenges of the research techniques, limitations of the study, and recommendations for future study will be discussed.

Challenges of each Methodology

The biggest limiting factor of content analysis methodology was the intensive amount of workload. As mentioned in chapter three, it took two human coders 40 hours to analyze 28 job announcements. Therefore, content analysis studies are limited to a relatively small sample size and become incompetent when dealing with vast amount of documents.

The second issue of content analysis is the subjective nature of coding scheme. As discussed in chapter three, two coders came up with different coding schemes under the same theoretical framework. Since humans have different perspectives, it is natural that their coding schemes will differ when opened coding is conducted. Therefore, choosing one coding scheme over another can be a challenge for content analysis researchers.

Text mining clearly has the advantage in its capability for processing large volumes of data at great speed. In this study, it took the computer less than five minutes to pre-process 3046 job announcements, and another five minutes to execute the LDA algorithm.

However, the efficiency of text mining does come with a price. First, it requires the researcher to develop knowledge of how text mining works, the capability to map research questions to text mining solutions, and the skill of using text mining software or creating customized tools. The process for developing the skillset has a steep learning curve and is a long-term endeavor. Secondly, in order to let computers automate the process with minimal human intervention, raw data typically needs to cleaned and pre-processed prior to data analysis. This is why a data preparation module was needed for the current study. Though there are existing tools that can execute standard text mining algorithms, the data cleaning and processing tasks typically require the creation of a customized tool or solution, which can be intellectually challenging and labor intensive.

Limitations of the Study

There are several limitations to this study. First, the research design for the content analysis was not optimized at its best. In fact, reasonable compromises were made for the sake of simplicity and efficiency. For example, the two coders could only afford two weeks for this study, and only 28 job documents were selected and analyzed. Given sufficient amount of time, interrater reliability could have been measured to check whether significant agreement was reached between the two coders. However, the study was focused on comparing the methodology more than the actual results. The goal of the study was to obtain some preliminary results for human coding so that a comparison between the two methodologies could be conducted. Therefore, acceptable compromises were made to save time for the human coders.

Secondly, this study only concerns one type of text mining application (topic modeling) and only uses one particular text mining software (T-LAB). Therefore, the research procedure may not be applicable to other text mining applications and analytical programs.

The third limitation is the usefulness of the results. The current dataset, with 67.6% from K12 and 27.5% from Higher Education, has a strong K12 focus. Therefore, it cannot represent all career opportunities in the field. The identified job categories and required competencies are domain-specific, and may not be applicable to non-educational settings such as business, industry, or government.

The last limitation applies to all text mining studies in general. Although the potential contribution of textual mining packages to the social studies is great, certain limitations should be noted. To date, there is no program that has the capacity to understand context-related meanings like a human reader. The interpretation of meaning must come from a human reader. This inability to understand texts in context is likely to remain a limitation of textual mining/analysis tools into the foreseeable future (Hoffman & Waisanen, 2015, p.179).

Textual analysis programs can perform sophisticated semantic analysis on a large scale, yet they are unable to understand meanings at the sentence-level syntax. For example, the following two sentences contain the same words but differ in syntax; computer programs are unable to detect the difference in meaning:

- I don't care how others think of me, but I care what you think of me.
- I don't care what you think of me, but I care how others think of me.

Knowing and acknowledging the limitations, as well as the potential capabilities, sets the foundation for how and when we should use computer-aids to accomplish our research goals.

Recommendations for Future Research

Many scholars have compared the human-based content analysis with the computer-assisted approach, including text mining. For example, Schnurr, Rosenberg, and Oxman (1993) conducted a study analyzing open-end speech, and found low agreement in results between the two procedures. Nacos et al. (1991) applied both human coding and computer-assisted approaches on political news coverage from the same data set and found satisfactory agreement between the two. They concluded that computer-based techniques should be treated as aids to complement content analysis, not as a replacement of highly advanced cognitive skills that only human coders possess. From a methodological perspective, Aureli (2017) stated that content analysis and text mining are, “not irreconcilable methods,” which seems to suggest there is a possibility that the two methods can work together.

Still, researchers haven't reached an agreement on the best practices for using computer tools in content analysis (Krippendorff, 2004). The evidence and findings in this study suggest that two techniques can complement each other, and it is possible to combine the two techniques while considering text mining to be an extension of the computer-aided content analysis.

A recommendation for future research concerning topic modeling is (a) to use topic modeling to identify themes and generate a report of meaningful context for each theme, and (b) to apply content analysis to code each theme manually. This combined approach can leverage the advantages of both techniques (the efficiency of text mining and the meaningfulness of content analysis), and thus enhance the quality and rigor of study.

References

AECT (Association for Educational Communications and Technology). (2012). *AECT Standards, 2012 version*.

Amadio, W. J., & Procaccino, J. D. (2016). Competitive analysis of online reviews using exploratory text mining. *Tourism and hospitality management, 22*(2), 193-210. doi:10.20867/thm.22.2.3

Ananiadou, S., & Mcnaught, J. (2006). *Text mining for biology and biomedicine*: Citeseer.

Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *International Journal of Digital Accounting Research*.

Barison, M., & Santos, E. (2011). The competencies of BIM specialists: a comparative analysis of the literature review and job ad descriptions. In *Computing in Civil Engineering (2011)* (pp. 594-602).

Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay*, 20-24.

Brill, J. M., Bishop, M. J., & Walker, A. E. (2006). The competencies and characteristics required of an effective project manager: A web-based Delphi study. *Etr&D-Educational Technology Research and Development, 54*(2), 115-140. doi:DOI 10.1007/s11423-006-8251-y

Bullough Sr, R. V., & Brumbaugh, W. D. (1974). The Development of a Competency-Based Model for Use in Instructional Technology.

Byun, H. (2000). *Identifying job types and competencies for instructional technologists: A five-year analysis.* (9993525 Ph.D.), Indiana University, Ann Arbor.

Carnevale, A. P., Jayasundera, T., & Repnikov, D. (2014). Understanding online job ads data.

Georgetown University, Center on Education and the Workforce, Technical Report (April).

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management, 54*(5), 807-817.

Drisko, J. W., & Maschi, T. (2015). *Content analysis*: Pocket Guides to Social Work R.

Durmaz, T. (2012). Competencies of instructional technologists graduated from computer education and instructional technology department as required by information technology firms. *Yayımlanmamış Yüksek Lisans Tezi, Ortadoğu Teknik Üniversitesi, Ankara*.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*: Cambridge University Press.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*: New York: McGraw-Hill Humanities/Social Sciences/Languages.

Gale, L. E., & Pol, G. (1975). Competence: A Definition and Conceptual Scheme. *Educational Technology, 15*(6), 19-25.

Gathercole, M., & Thurairajah, N. (2014). *The influence of BIM on the responsibilities and skills of a project delivery team.* Paper presented at the International Conference on Construction in a Changing World, Heritance Kandalama, Sri Lanka.

Gaur, A., & Kumar, M. (2018). A systematic approach to conducting review studies: An assessment of content analysis in 25 years of IB research. *Journal of World Business, 53*(2), 280-289. doi:10.1016/j.jwb.2017.11.003

Greene, D., O'Callaghan, D., & Cunningham, P. (2014). *How many topics? stability analysis for topic models.* Paper presented at the Joint European Conference on Machine Learning and

Knowledge Discovery in Databases.

Hoffman, D., & Waisanen, D. (2015). At the digital frontier of rhetoric studies: An overview of tools and methods for computer-aided textual analysis. *Rhetoric and the digital humanities*, 169-183.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research, 15*(9), 1277-1288.

Huggins, A. C., Ritzhaupt, A. D., & Dawson, K. (2014). Measuring information and communication technology literacy using a performance assessment: Validation of the student tool for technology literacy (ST2L). *Computers & Education, 77,* 1–12.

IBSTPI (2017). International board of standards for training performance and instruction. Available at: http://ibstpi.org/

Ignatow, G., & Mihalcea, R. (2017). *An introduction to text mining: Research design, data collection, and analysis*. Sage Publications.

ISTE (2019). ISTE standards. Retrieved from https://www.iste.org/standards/standards.

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism, 4*(1), 89-106.

Januszewski, A., & Molenda, M. (2007). Educational technology: A definition with commentary. New York: Taylor & Francis Group.

Jiang, F., & McComas, W. F. (2014). Analysis of Nature of Science Included in Recent Popular Writing Using Text Mining Techniques. *Science & Education, 23*(9), 1785-1809. doi:10.1007/s11191-014-9703-0

Kang, Y., & Ritzhaupt, A. D. (2015). A Job Announcement Analysis of Educational Technology Professional Positions: Knowledge, Skills, and Abilities. *Journal of Educational*

Technology Systems, 43(3), 231-256. doi:10.1177/0047239515570572

Kimmons, R., & Veletsianos, G. (2018). Public Internet Data Mining Methods in Instructional Design, Educational Technology, and Online Learning Research. *TechTrends, 62*(5), 492-500. doi:10.1007/s11528-018-0307-4

Krippendorff, K. (2004). Content analysis: An introduction to its methodology Thousand Oaks. *Calif.: Sage*.

Ley, T., & Albert, D. (2003). Identifying employee competencies in dynamic work domains: methodological considerations and a case study. *J. UCS, 9*(12), 1500-1518.

Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education, 52*(2), 481-495. doi:10.1016/j.compedu.2008.10.005

Lu, C., & Miller, L. (2002). Instructional technology competencies perceived as needed by vocational teachers in Ohio and Taiwan. *Journal of Vocational Education Research, 27*(3), 319-329.

Maer-Matei, M. M., Mocanu, C., Zamfir, A.-M., & Georgescu, T. M. (2019). Skill Needs for Early Career Researchers—A Text Mining Approach. *Sustainability, 11*(10), 2789.

Marcial, D. E., & Rendal, J. B. (2014). Are Teachers Technophobic? Measuring Information and Communication Technology Competency among Teacher Educators in Central Visayas, Philippines. *Malaysian Journal of Distance Education*, *16*(2), 1-17.

Marrelli, A. F. (1998). An introduction to competency analysis and modeling. *Performance Improvement, 37*(5), 8-17.

May, J. D. (1980). AN ANALYSIS OF MEDIA MANAGEMENT: DETERMINING THE MANAGERIAL ACTIVITIES AND FUNCTIONAL ROLES OF THE UNIVERSITY

MEDIA SERVICES DIRECTOR THROUGH STRUCTURED OBSERVATION.

Competency. (2020). In *Merriam-Webster.com dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/competency

Moallem, M. (1995). Analysis of Job Announcements and the Required Competencies for Instructional Technology Professionals.

Molenda, M., & Pershing, J. (1992). *Corporate instructional designers curriculum needs assessment.* Paper presented at the National Society for Performance Improvement National Conference.

Morlan, J. E., & Lu, M.-Y. (1993). A Survey of Media and Instructional Technology Competencies Needed by Business, Industry, Health Professions, Agencies, Military Trainers, and Independent Contractors in Northern California, USA.

Mullins, L. (1985). Job analysis — know the job and the person to do it. *International Journal of Hospitality Management, 4*(4), 181-183. doi:https://doi.org/10.1016/0278-4319(85)90057-X

Nacos, B. L., Shapiro, R. Y., Young, J. T., Fan, D. P., & Kjellstrand, T. (1991). Content analysis of news reports: Comparing human coding and a computer-assisted method. *Communication (New York), 12*(2), 111-128.

Neuendorf, K. (2010). A.(2002) The content analysis guidebook. *Thousand Oaks, CA*.

Nordquist, R. (2015). Glossary of grammatical and rhetorical terms. Retrieved from https://www.thoughtco.com/head-words-tern-1690922.

Northrup, P. T. (1997). Instructional Technology Benchmarks for Teacher Preparation Programs and K-12 School Districts.

Osorio, N. L. (1999). An analysis of science-engineering academic library positions in the last three decades. *Issues in Science and Technology Librarianship, 24*.

Parry, S. B. (1996). The quest for competencies. *Training, 33*(7), 48.

Patterson, A. C. (1985). Preparing Educational Technologists. *Training and Development Journal, 39*(12), 38-39.

Richey, R. C., Fields, D. C., & Foxon, M. (2001). *Instructional design competencies: The standards*: ERIC.

Ritzhaupt, A., Martin, F., & Daniels, K. (2010). Multimedia competencies for an educational technologist: A survey of professionals and job announcement analysis. *Journal of Educational Multimedia and Hypermedia, 19*(4), 421-449.

Ritzhaupt, A. D., & Kumar, S. (2015). Knowledge and Skills Needed by Instructional Designers in Higher Education. *Performance Improvement Quarterly, 28*(3), 51-69. doi:10.1002/piq.21196

Ritzhaupt, A. D., Martin, F., Pastore, R., & Kang, Y. (2018). Development and validation of the educational technologist competencies survey (ETCS): knowledge, skills, and abilities. *Journal of Computing in Higher Education, 30*(1), 3-33. doi:10.1007/s12528-017-9163-z

Sampson, D., & Fytros, D. (2008). Competence models in technology-enhanced competence-based learning. In *Handbook on information technologies for education and training* (pp. 155-177): Springer.

Sanchez, J. I. (1994). From documentation to innovation: Reshaping job analysis to meet emerging business needs. *Human Resource Management Review, 4*(1), 51-74. doi:https://doi.org/10.1016/1053-4822(94)90004-3

Schnurr, P. P., Rosenberg, S. D., & Oxman, T. E. (1992). Comparison of TAT and free speech techniques for eliciting source material in computerized content analysis. *Journal of* *personality assessment, 58*(2), 311-325.

Shank, J. D. (2006). The blended librarian: A job announcement analysis of the newly emerging position of instructional design librarian. *College & Research Libraries, 67*(6), 514-524.

Shekhar, A., & Venkatesan, M. (2018). *A Bag-of-Phonetic-Codes Modelfor Cyber-Bullying Detection in Twitter.* Paper presented at the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT).

Short, J. C., & Palmer, T. B. (2008). The application of DICTION to content analysis research in strategic management. *Organizational Research Methods, 11*(4), 727-752.

Singh, P. (2008). Job analysis for a changing workplace. *Human Resource Management Review, 18*(2), 87-99.

Spector, J. M. (2015). *The SAGE encyclopedia of educational technology*. Sage Publications.

Spitzer, D. R. (1988). Instructional/performance technology competencies. *Performance & Instruction*.

Stolovitch, H. D., Keeps, E. J., & Rodrigue, D. (1995). Skill sets for the human performance technologist. *Performance Improvement Quarterly, 8*(2), 40-67.

Succar, B., Sher, W., & Williams, A. (2013). An integrated approach to BIM competency assessment, acquisition and application. *Automation in Construction, 35*, 174-189. doi:https://doi.org/10.1016/j.autcon.2013.05.016

Sugar, W., Brown, A., Cafeteria, O., & Daniels, L. (2007). Media Production Curriculum and Competencies: Identifying Entry-level Multimedia Production Competencies and Skills of Instructional Design and Technology Professionals: Results from a Biennial Survey. *2007 Annual Proceedings-Anaheim: Volume*, 284.

Sugar, W., Hoard, B., Brown, A., & Daniels, L. (2012). Identifying multimedia production competencies and skills of instructional design and technology professionals: An analysis of recent job postings. *Journal of Educational Technology Systems, 40*(3), 227-249.

Surry, D. W., & Robinson, M. A. (2001). A Taxonomy of Instructional Technology Service Positions in Higher Education. *Innovations in Education and Teaching International, 38*(3), 231-238. doi:10.1080/14703290110051406

Tennyson, R. D. (2001). Defining core competencies of an instructional technologist. *Computers in Human Behavior, 17*(4), 355-361. doi:https://doi.org/10.1016/S0747-5632(01)00014-0

T-LAB. (2019). User's Manual T-LAB Plus 2019. Retrieved from https://www.tlab.it/

Uhm, M., Lee, G., & Jeon, B. (2017). An analysis of BIM jobs and competencies based on the use of terms in the industry. *Automation in Construction, 81*, 67-98. doi:10.1016/j.autcon.2017.06.002

Weber, R. P. (1990). *Basic content analysis*. Sage.

Wowczko, I. (2015). *Skills and vacancy analysis with data mining techniques.* Paper presented at the Informatics.

Yang, Q., Zhang, X., Du, X., Bielefield, A., & Liu, Y. (2016). Current market demand for core competencies of librarianship—A text mining study of American Library Association's advertisements from 2009 through 2014. *Applied Sciences, 6*(2), 48.

Zhai, C., & Massung, S. (2016). *Text data management and analysis: a practical introduction to information retrieval and text mining*: Association for Computing Machinery and Morgan & Claypool.

Appendix A — Abridged Version of Human Coding Results (Coding Task Three)

No.	Doc	Category	Segment	Code	Coder
1	810	c10	In addition to the specific skills described below, a successful candidate will be a patient problem solver, observational, have excellent written and oral communication skills, and will be self-directed and highly organized.	S1, S6, S7, S8, O1	M
2	810	c10	Provide technology expertise, current research, and best practice benchmarks to design, measure, and evaluate technology integration, Knowledge of and experience applying the ISTE standards for student and teacher technology integration;	k14, A16, A17, A4.1	M
3	810	c10	EXPERIENCE / QUALIFICATIONS & SKILLS: B.A.	O	M
4	810	c10	Degree in a related field Proven experience in instructional design, technology integration, working with teachers and teaching students;	K3, K14, A2, A3	M
5	810	c10	Highly proficient in Google Apps for Education, experience using and managing Moodle or equivalent (LMS);	k1.1, A19	M
6	810	c10	Experience with multiple platforms including iOS, Windows 10, and Mac OSX	K9	M
7	810	c10	Candidates must have a desire to contribute to the education of young women as well as the ability to build positive, constructive relationships with students.	A3	M
8	810	c10	The candidate will be joining a school with a collaborative faculty working to build a hands-on, relevant curriculum that delivers knowledge that sticks through innovative programming for girls in the STEAM disciplines.	A9, A4	M
9	810	c10	Ideal candidates will be interested in preparing our students to live and work in a global and multicultural society.	O	M
10	1637	c10	In addition to the specific skills described below, a successful candidate will be a patient problem solver, observational, have excellent written and oral communication skills, and will be self-directed and highly organized.	S1; S6; O1	Q

369	5698	c6	The ideal candidate will understand how to create real-world learning experiences and foster a design mentality and spirit of innovation within a progressive learning environment.	O3, A8	M
370	5698	c6	The strongest candidates will have experience working with faculty members and students in a diverse learning environment that features an interdisciplinary and project-based curriculum.	A3, A4	M
371	5698	c6	We are seeking a Maker Lab Teacher whose experience, talent, and energy will inspire and equip students with not only skills and proficiencies, but also a deep sense of the designed world and their own potential as creative, real-world designers, makers, and problem-solvers.		M
372	5698	c6	Develop and articulate a continuum of Maker skills, experiences and mindsets across grades K-8.	A9	M
373	5698	c6	Skills and Qualities		M
374	5698	c6	Experience across multiple fabrication domains, with a steadfast focus on safety.	K19.1	M
375	5698	c6	Ability to develop, build, define, stock, and organize lab space.	A5	M
376	5698	c6	Experience developing and implementing safety, maintenance, and tracking systems.	A7.5	M
377	5698	c6	sense of humor and strong communication skills.	S1, S5	M
378	5698	c6	Ability to connect with students who may have difficulties engaging in other academic areas.	a50	M
379	5698	c6	The desire and ability to contribute to student and community life inside and outside the classroom.	O	M
380	5698	c6	BA and prior teaching experience required.	A2	M
381	5698	c6	Demonstrated ability to design and deliver meaningful Maker experiences across K-8 grade levels.	a5	M
382	5698	c6	Teaching Credential / Master s degree preferred.		M

Appendix B — Researcher's Interpretation of Text Mining Topics

Machine Label	Topical Words (with Transformed Probability Value)	Interpretation	Decision
Adobe	Adobe (10)Captivate (10)Cut (10)Dreamweaver (10)Excel (10)illustrator (10)iMovie (10)Indesign (10)Photoshop (10)Powerpoint (10)	Software	Keep
Application	Attachment (10)Click (10)Commission (10)Electronically (10)Letter (10)Letters (10)Line (10)Name (10)Packet (10)Resume (10)	How to Apply	Discard
Candidate	Genuine (10)Sense (10)Can-Do (9)Compassion (9)Humor (9)High_Energy (8)Passionate (8)Self-Directed (8)Adolescent (7)Adept (7)	Personality	Keep
Certification	Acsi (10)Cissp (10)Driver (10)Georgia (10)Licensure (10)Pmp (10)Valid (10)Certify (9)Company (9)Credential (9)	Licensure Related	Discard
Check	Abuse (10)Check (10)Clearance (10)Contingent (10)Criminal (10)Csu (10)Fbi (10)Fingerprint (10)Immigration (10)Legally (10)	Background Check	Discard
Combination	Concentration (6)Combination (5)Comparable (5)Equivalent (4)Industrial (4)Consult (3)Entrepreneurship (3)Internship (3)One_Year (3)Qualify (3)	Uninterpretable	Discard
Committee	Advisor (10)Chair (10)Chaperone (10)Extracurricular (10)Involvement (10)Profession (10)Trip (10)Athletic (9)Dean (9)Serve (9)	Service Related	Discard
Communication	Time-Management (10)Prioritization (9)Attention (8)Multi-Tasking (8)Proofreading (7)Decision_Making (6)Detail (6)Oral (6)Organizational (6)Verbal (6)	Soft Skills	Keep
Data	Life_Cycle (10)Oracle (10)Peoplesoft (10)Query (10)Scrum (10)Statistical (10)Visualization (10)Crm (9)Iso (9)Nist (9)	Data Management And Analysis	Keep
Degree	Phd (10)Stem-Related (10)Masters (9)Second (9)Undergraduate (9)Doctorate (6)Advanced (5)Coursework (5)Doctoral (5)Graduate (5)	Education Requirement	Discard
Diversity	Anza (10)Appreciate (10)Cultural (10)Distinctive (10)Equity (10)Inclusion (10)Mutual (10)Population (10)Sensitivity (10)Socioeconomic (10)	Cultural Values	Keep

Duty	Construe (10)Exhaustive (10)Gathering (10)Intend (10)Listing (10)Classify (8)Summary (8)Duty (7)Responsibility (7)Assign (6)	Uninterpretable	Discard
Education	Ellucian (10)Librarianship (10)Professor (10)Promotion (10)Rank (10)Tenure (10)Autonomy (9)Banner (8)Clinical (7)Registrar (7)	Uninterpretable	Discard
Equivalent	Ged (10)Substitution (10)Substitute (9)Diploma (7)Specialization (7)Paid (6)Technology-Based (6)Combination (5)Relevant (5)Additional (4)	Uninterpretable	Discard
Essential	Angular (10)Bootstrap (10)Css3 (10)Html5 (10)Jquery (10)Ui (10)Back-End (9)Satisfactorily (9)Utilization (9)Ux (9)	Web Development	Discard
Hardware	Audio-Visual (10)Audiovisual (10)Cable (10)Install (10)Smartboards (10)Tv (10)Projector (9)Hardware (8)Repair (8)Scanner (8)	Hardware	Keep
Hour	Fraction (10)Holidays (10)Hour (10)Monday (10)Night (10)Occasional (10)Travel (10)Week (10)Weekend (10)Availability (9)	Work Hours And Travel Requirement	Discard
Library	Circulation (10)Collection (10)Landscape (10)Marc (10)Metadata (10)Catalog (9)Preservation (9)Repository (8)Vast (8)Literature (7)	Library And Information Science	Keep
LMS	Blackboard (10)Elearning (10)Face-To-Face (10)Hybrid (10)In-Person (10)LMS (10)Universal (10)Canvas (9)Moodle (9)Blended (8)	Learning Management System	Keep
Management	Progressively (10)Encompass (9)Increasingly (8)Managerial (8)Responsible (8)Supervisory (7)Hands_On (5)Oversight (5)Progressive (5)Budgetary (4)	Administrative Responsibility	Discard
Market	Analytics (10)Advertise (10)Brand (10)Calendar (10)Campaign (10)Engine (10)Facebook (10)Finalsite (10)Layout (10)Market (10)	Marketing	Keep
Mission	Biblical (10)Catholic (10)Christ (10)Christian (10)Episcopal (10)Faith (10)God (10)Jesus (10)Lifestyle (10)Lives (10)	Religious Belief	Keep
Multiple	Interrelated (10)Interruption (10)Juggle (10)Prioritize (10)Simultaneous (10)Simultaneously (10)Strict (10)Tight (10)Timeline (10)Confidential (9)	Uninterpretable	Discard

Parent	Alumnus (10)Orally (10)Punctuation (10)Verbally (10)Concise (9)Grammar (9)Spelling (9)Productively (8)Cooperatively (7)Disposition (7)	Uninterpretable	Discard
Plan	Agent (10)Consistency (10)Fiscal (10)Ks (10)Long-Term (10)Prioritizing (10)Retain (10)Attract (9)Lively (9)Adoption (8)	Uninterpretable	Discard
Pounds	Arms (10)Bend (10)Carry (10)Carrying (10)Climb (10)Comfortably (10)Crawl (10)Crouch (10)Dexterity (10)Finger (10)	Physical Requirement	Keep
Problem	Cause (10)Collect (10)Diagnose (10)Gather (10)Investigate (10)Logical (10)Mathematical (10)Problem (10)Resolution (10)Solve (10)	Problem Solving	Keep
Production	Audio (10)Cnc (10)Commonly (10)Cutter (10)Film (10)Laser (10)Machinery (10)Prototyping (10)Recording (10)Shop (10)	Design And Production	Keep
Range	Array (10)Spectrum (9)Range (8)Resilienc (8)Old (7)Wide (7)Age_Group (6)Broad (6)Learn (6)Different (5)	Uninterpretable	Discard
Relationship	Co-Worker (10)Credibility (10)Non-Technical (10)Rapport (9)Trust (9)Courteous (8)Initiate (8)Relationship (8)Stakeholder (8)Sustain (8)	Teamwork	Keep
Report	Cards (10)Compile (10)Seminar (10)Trainings (10)Computerize (9)Acquire (8)Accurate (8)Attendance (8)Discussion (7)Vocational (7)	Uninterpretable	Discard
Robotic	Arduino (10)Asp (10)C# (10)Competition (10)Ev3 (10)Frc (10)Java (10)Lego (10)Mechanical (10)Object-Oriented (10)	Robotics And Coding	Keep
School	Childhood (10)Elementary-Aged (10)K-5 (10)Spanish (10)Mentoring (9)Bilingual (7)Boy (7)Early (7)Elementary (7)French (7)	Uninterpretable	Discard
Science	Pre-K (10)Cis (9)Elective (9)Introductory (9)Keyboarding (8)Placement (8)Math (7)Mathematic (7)Teach (7)Ap (6)	Uninterpretable	Discard
Security	Ada (10)Audit (10)Cyber (10)Cybersecurity (10)Detection (10)Fair (10)Ferpa (10)Hipaa (10)Intrusion (10)Privacy (10)	Privacy And Security	Keep
Service	Accommodation (10)Provider (10)Reasonable (10)Seamless (10)Capital (9)Servicenow (9)End_Users (8)Interfacing (8)Itsm (8)Architect (7)	Uninterpretable	Discard

Skill	Superb (8)Superior (8)Exemplary (7)Collegiality (6)Insight (6)Interpersonal (6)Problem-Solving (6)Analytic (5)Couple (5)Customer (5)	Too Broad	Discard
Standard	Ambitious (10)Benchmark (10)Iste (10)Adopt (8)Scholar (8)Washington (8)Cognitive (7)Stay (7)Target (7)Girl (6)	Uninterpretable	Discard
Statu	Benefit (10)Discrimination (10)Employer (10)Equal (10)Ethnicity (10)Expression (10)Hall (10)Ix (10)Marital (10)Military (10)	Uninterpretable	Discard
Subject	Option (10)Universityvalid (10)Career (9)Instruct (9)Wage-Earning (8)Approve (7)Endorsement (7)Internship (7)Match (7)Society (7)	Uninterpretable	Discard
Support	Informal (10)Day-To-Day (9)Formal (9)School-Based (7)High_Quality (6)High-Quality (6)Publish (6)Central (5)Desirable (5)Tech (5)	Uninterpretable	Discard
Teacher	Enrichment (10)Sequence (10)Developmentally (9)Introduce (9)Powerful (9)Citizenship (8)Hub (8)Inside (8)Strengthen (8)Embed (7)	Uninterpretable	Discard
Teaching	Compensation (10)Courtesy (10)Full_Time (10)Ib (10)Myp (10)Negotiable (10)Salary (10)Nationally (9)Not-For-Profit (9)Programme (9)	Uninterpretable	Discard
Theme_00	Ala-Accredited (10)Ala (10)Four-Year (10)Mis (10)Mlis (10)Regionally (10)Accredit (9)Mls (9)Graduation (8)Liberal (8)	Uninterpretable	Discard
Theory	Modality (10)Standards-Based (10)Theory (10)Differentiation (8)Pedagogical (8)Bc (7)Contemporary (7)Mode (7)Method (6)Pedagogy (6)	Instructional Theory	Keep
Think	Inquiry-Based (10)Personalize (10)Real-World (10)Computational (9)Emphasize (9)Engaged (9)Experiential (9)Problem-Based (9)Purposeful (9)Connection (8)	Thinking Mindset	Keep
Vision	Adjust (10)Acuity (10)Breadth (10)Conversation (10)Correspondence (10)Depth (10)Diagram (10)Failure (10)Perception (10)Read (10)	Uninterpretable	Discard
Windows	android (10)Chrome (10)Cisco (10)Configuring (10)Dhcp (10)Directory (10)Dns (10)Filemaker (10)Filter (10)Git (10)		

Work	Adaptable (10)Calm (10)Pace (10)Player (10)Team-Oriented (10)Fast (9)Fun (9)Stressful (9)Self-Motivated (8)Accountable (7)	Work Style	Keep
Year	Commercialization (9)Limitation (9)Cloud-Based (6)Coordination (6)One_Year (5)Alma (4)Corporate (4)Programmer (4)Trainer (4)Workday (4)	Uninterpretable	Discard

Appendix C — Headwords vs. Areas of Expertise

	Administrator	Advisor	Advocate	Analyst	Architect	Assistant	Associate	Chair	Chancellor	CIO
Web Development, Management	x			x		x				
LMS	x			x						
Instructional Design/Educational Technology	x	x	x	x	x	x	x		x	
Library Information Science				x		x				
Multimedia/Digital media						x	x	x		
Technology-infused spaces						x				
Information Technology/Computer Science				x			x	x		x
Marketing and Communications							x			
STEM/STEAM						x		x		
Distance/Online/E-Learning										
Research										
Assistive Technology						x				
Social media										
Project Management										
Leadership										

Educational Information Management	x					x	x			

	Coach	Collaborator	Consultant	Coordinator	Dean	Designer	Developer	Director	Editor	Educator
Web Development, Management			x	x	x		x	x	x	
LMS										
Instructional Design/Educational Technology	x		x	x	x	x		x		x
Library Information Science				x	x			x		
Multimedia/Digital media	x			x	x		x	x	x	x
Technology-infused spaces		x		x				x		x
Information Technology/Computer Science	x			x		x	x	x		
Marketing and Communications				x				x		
STEM/STEAM	x			x	x			x		x
Distance/Online/E-Learning				x		x		x		
Research				x				x		
Assistive Technology				x				x		
Social media				x						
Project Management								x		
Leadership								x		

Educational Information Management	x

	Engineer	Facilitator	Faculty	Fellow	Head	Instructor	Integrationist	Integrator	Intern	Leader
Web Development, Management										
LMS										
Instructional Design/Educational Technology	x	x	x	x	x	x	x	x	x	x
Library Information Science					x	x	x			
Multimedia/Digital media	x		x		x	x		x		
Technology-infused spaces	x	x	x	x		x			x	
Information Technology/Computer Science	x		x		x	x	x	x	x	
Marketing and Communications					x	x				
STEM/STEAM			x			x		x	x	x
Distance/Online/E-Learning										
Research										
Assistive Technology						x				
Social media										
Project Management										x
Leadership										

Educational Information Management

	Lecturer	Liaison	Librarian	Manager	Mentor	Officer	Operator	President	Producer	Professor
Web Development, Management			x	x						
LMS				x						
Instructional Design/Educational Technology	x	x	x	x	x	x		x		x
Library Information Science			x	x		x		x		
Multimedia/Digital media			x	x			x		x	
Technology-infused spaces				x	x					
Information Technology/Computer Science				x		x		x		
Marketing and Communications				x		x				
STEM/STEAM			x							
Distance/Online/E-Learning			x	x						
Research				x				x		
Assistive Technology										
Social media				x						
Project Management				x						
Leadership				x						

Educational Information Management	x

	Provost	Registrar	Specialist	Steward	Strategist	Supervisor	Support	Teacher	Technician
Web Development, Management			x						
LMS							x		
Instructional Design/Educational Technology			x		x	x	x	x	x
Library Information Science			x						x
Multimedia/Digital media			x				x	x	x
Technology-infused spaces			x			x	x	x	x
Information Technology/Computer Science	x		x					x	x
Marketing and Communications			x					x	
STEM/STEAM			x				x	x	
Distance/Online/E-Learning			x						
Research								x	
Assistive Technology			x				x		x
Social media			x		x				
Project Management	x		x						
Leadership			x						

Educational Information Management	x	x	x

	Technologist	Tinkerer	Trainer	Webmaster
Web Development, Management	x			x
LMS				
Instructional Design/Educational Technology	x	x	x	
Library Information Science				
Multimedia/Digital media				
Technology-infused spaces		x		
Information Technology/Computer Science				
Marketing and Communications				x
STEM/STEAM				
Distance/Online/E-Learning				
Research				
Assistive Technology				
Social media				
Project Management				
Leadership				

Educational Information Management