Developing an Intelligent Assistant for the Audit Plan Brainstorming Session

by

Qiao Li

A dissertation submitted to the

Graduate School - Newark

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Professor Miklos Vasarhelyi

and approved by

Dr. Miklos A. Vasarhelyi

Dr. Alexander Kogan

Dr. Helen Brown-Liburd

Dr. Donald Warren

Newark, New Jersey

January 2019

ABSTRACT OF THE DISSERTATION

Developing an Intelligent Assistant for the Audit Plan Brainstorming Session

By QIAO LI

Dissertation Director:

Dr. Miklos A. Vasarhelyi

During the initial stages of an audit, audit engagement teams are required to conduct brainstorming sessions to evaluate risk factors and discuss the susceptibility of the entity's financial statements to material misstatement, either as a result of error or fraud (AICPA 2012). At present, the most commonly used decision support tool for audit plan brainstorming is the checklist (Bellovary and Johnstone 2007), which has shown limitations. Large audit firms have been investing in substantive resources to the utilization of Artificial Intelligence (AI) to take advantage of their past audit experience and industry knowledge (Kokina and Davenport 2017). This dissertation suggests that the latest intelligent assistant technology can be applied in the auditing domain to provide risk assessment decision supports to audit engagement teams during audit plan brainstorming discussions.

Firstly, an interactive audit cognitive assistant framework is proposed to provide auditors with information retrieval and risk assessment help. The proposed framework provides a new method of knowledge organization for the audit domain, which potentially develops a knowledge base that stores many auditors' knowledge and experience in audit risk identification and assessment.

Furthermore, a Natural Language Processing (NLP)-based audit plan knowledge discovery system (APKDS) is proposed. By applying NLP techniques, the proposed system can continuously collect auditors' professional experience and expertise in audit brainstorming discussions and transfer the discussion into classified knowledge for future use. The output of the proposed system can provide insights into how auditors identify and assess risks during the audit plan and how audit decisions are made.

Finally, we propose a prototype for the APKDS framework and illustrate the development of important modules in the system. Experimental brainstorming meeting recordings are used as training and testing datasets in model building and training. We demonstrate that the proposed objectives of the system can be realized and the proposed system can provide effective decision-making support to auditors. In the end, we proposed the potential application of the audit cognitive assistant to other audit phases.

ACKNOWLEDGEMENTS

It is a sentimental moment to write this note of thanks as the finishing touch of my dissertation. This thesis represents the cohesion of painstaking efforts of several years of work, and could not have been accomplished without the love and support from many people. Here I would like to express my greatest appreciation to all the people who provided me tremendous support and help during my Ph.D. study.

I would like first to express my deep gratitude to my mentor and adviser, Prof. Miklos A. Vasarhelyi. He brought me to academia and encouraged me to explore on my own. He provides me intellectual guidance, inspiring insights, inexhaustible patience, and extraordinary knowledge, which are necessary to survive and thrive the graduate school and the beyond. When my step faltered, he often told me the quote “if we knew what it was we were doing, it would not be called research, would it?” (by Albert Einstein). I thank him for generously giving me motivation, support, time, understandings, assistance, opportunities, trust, and friendship for all these years. Moreover, I want to thank him for providently positioning and guiding me on the right direction of the pioneer research work of audit data analytics. He helped me to be a better writer, speaker, and scholar.

I also sincerely grateful to those who served in my committees: Prof. Alexander Kogan, Prof. Helen Brown-Liburd and Prof. Donald Warren. I deeply appreciate their time and efforts in reviewing my dissertation and attending my dissertation defense. All of them not only provide thought-provoking and constructive suggestions on my research, but also offer tremendous support and help in my academic job-hunting process. I would like to acknowledge my gratitude to Prof. Dan Palmon for his consistent support as department chair. Special thanks are due to Prof. Won Gyun No, Prof. Hussein Issa, Prof. Michel Alles, Prof. Kyungha Lee, Prof. Kevin C. Moffitt and Prof. Irfan Bora in the AIS department of Rutgers Business School and for their help with both my job search and research work.

I would also like to acknowledge the CARLAB (Continuous Auditing & Reporting Lab) of Rutgers Business School for supplying me with the best equipment and working environment that helped me to accomplish much of this work. It is an honor to be affiliated with such a great research group and a loving family. I also owe a hefty amount of thanks to my friends and colleagues at Rutgers: Dr. Victoria Chiu, Dr. Jun Dai, Dr. Kyunghee Yoon, Xuan Peng, Dr. Tiffany Chiu, Feiqi Huang, Kexing Ding, Yue Liu, Dr. Ting Sun, Dr. Jingyuan Yang, Dr. Meng Qu, Dr. Meng Li et al. for their help, friendship, and valuable suggestions.

Finally, my deepest thanks and love to my husband Junming Liu, who gave me infinite joy, love, and support to make this work possible. Especially, I thank him for providing me tremendous professional suggestions and help to my research work. I also thank him for making it very challenging to take care of him in daily life, which helped me to be a stronger woman and have the courage to face other difficulties such as completing the thesis. I owe my deepest gratitude and love to my parents and in-laws for their unwearied devotion and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Audit tools have been widely applied in audit firms to provide auditors assistance in a variety of tasks. With technological advances, audit firms have been able to increase the level of decision support imbedded within the firms' audit support tools (Dowling et al. 2008). There has been a trend that of audit firms start to utilize Artificial Intelligence (AI) to develop new audit tools to provide advanced data analytics and audit decision support (Issa et al. 2016; Kokina and Davenport 2017). AI-enabled technology can mimic the "cognitive" functions of humans, such as "learning" and "problem solving" (Russell and Norvig 2009). Some main problems that AI research tries to solve include natural language processing, reasoning, knowledge representation, planning, learning, etc. (Russell and Norvig 2003; Luger 2005).

Applications of artificial intelligence are particularly suitable for the audit profession because it has been increasingly challenging for auditors to analyze large volumes of structured and unstructured data and then identify potential risks and make various audit decisions (Kokina and Davenport 2017). The learning and analytical capabilities of AI-enabled tools make them able to understand users' behaviors and interests through the human-computer interactions and then provide decision supports such as answers to questions and recommendations to users.

One example of emerging AI-enabled tools is the Cognitive Assistant, or Intelligent Personal Assistant (IPA), which has become one of the fundamental devices in mobile operating systems (Canbek and Mutlu 2016). Popular commercial cognitive assistants include Apple's Siri, Google Now, Microsoft Cortana, Amazon's Echo/Alexa, IBM's Watson etc. (Canbek and Mutlu 2016; Mehrez et al. 2013; Ebling 2016; Bellegarda 2013;

Strayer et al. 2017). Cognitive Assistants interact with users using natural language and provide instant assistance to users through answering questions, providing recommendations and performing actions (Hauswald et al. 2015). Current IPAs have shown their usefulness in people's daily lives such as searching for relevant information, selecting goods and placing orders and scheduling events, but the technology has been slow to be applied to the work environment. They have great potential to be used for business situations where information retrieval and decision-making support are needed.

In an audit plan brainstorming risk assessment discussion, auditors need to use their professional judgement and experience to extract most important information from various financial and non-financial data and then identify risks and make audit decisions. If there are effective audit tools that can provide in-time decision aids to an audit engagement team such as locating and extracting relevant information and recommend possible risk areas, then auditors can focus on areas that require higher-level judgment (Kokina and Davenport 2017). Towards this direction, this thesis discusses the potential application of cognitive assistant technology in the auditing domain and provides a vision of future audit tools. This thesis proposes an interactive intelligent cognitive system that can be used in the audit brainstorming sessions to help auditors evaluate information and make subsequent judgments in risk discovery and risk assessment.

This chapter introduces the motivation and method of this thesis and provides a literature review of the related concepts.

1.1 Background & Preliminaries

In this section, we provide a brief overview of the context and preliminaries in audit brainstorming sessions and cognitive assistants.

1.1.1 Audit brainstorming sessions

Brainstorming sessions are mandated in two auditing standards, SAS No. 99 (AICPA 2002), Consideration of Fraud in a Financial Statement Audit, and SAS No. 109, Understanding the Entity and Its Environment and Assessing the Risk of Material Misstatement. Audit brainstorming meetings allow engagement teams to identify risks and discuss how a material misstatement, whether fraudulent or erroneous, could occur (PCAOB 2010; Landis et al. 2008). Auditors are expected to generate effective audit procedures for detecting risks after the discussion. The most commonly procedure for brainstorming is using a checklist in an open-ended form (Bellovary and Johnstone 2007).

Brainstorming meetings are very important for three main reasons. Firstly, identifying potential fraud areas and making an audit plan in the brainstorming session have great impact on auditors' subsequent audit performance (Carpenter 2007). Secondly, studies showed that team discussions in the brainstorming sessions help auditors generate more high-quality ideas on risks compared to auditors' individual assessment. Thirdly, auditors' discussions in brainstorming meetings provide valuable knowledge that worth collecting and extracting. The meeting involves various topics on different risks and how different engagement teams evaluate information, identify and assess risks and make decisions. The contents integrate various information such as documents, auditors' experience, innovative ideas, etc., which can be used as decision aids for future audit engagement cases if collected. Therefore, this thesis focuses on two main objectives: one is to propose a new type of decision support tool which aims at helping auditors identify and assess risks during brainstorming meetings, and another is to collect and extract important risk assessment knowledge from audit brainstorming conversations and prepare it in a machine-readable format for future use.

1.1.2 Cognitive assistant

Cognitive Assistants provide instant assistance to users by answering users' questions, making recommendations, and performing users' commands (Hauswald et al. 2015). Since humans have been increasingly getting used to expressing their general needs to computer systems and let the system to help them in a stochastically consistent manner, computer tools like intelligent personal assistants will be more commonly accepted and applied in different domains (Bellegarda 2013).

Cognitive Assistants can provide solutions for both simple tasks and complicated intelligent/cognitive tasks. Examples of simple tasks that most IPAs can complete include setting a timer, sending messages, getting directions, creating reminders, reminding daily appointments, responding to questions, invoking apps, finding the items on online store and adding them to shopping cart etc. (Canbek and Mutlu 2016). For example, Siri could send messages by integrating with default IOS functional applications of contacts and text message, and it could answer users' questions with the help of web searches from Google, Bing, Yahoo, Apple Maps etc. IMB Watson, developed with complicated cognitive computing technologies, could support users to make more informed decisions.

At present, the most important benefits of cognitive assistants are retrieving information and executing users' commands through its built-in applications. For example, researchers from University of Michigan (Hauswald et al. 2015) built an open intelligent personal assistant system Sirius (Figure 1), which works similar to other IPAs such as Siri and Microsoft Cortana in accepting voice commands. The only difference is that it can also process image commands and allow users to ask questions about what they are seeing. Although the image processing feature in Sirius is not needed in our study, we can find that the most important components in current IPAs include Natural Language Processing (NLP) modules and the Question Answering (QA) System (Hearst 2011; Mehrez et al. 2013).

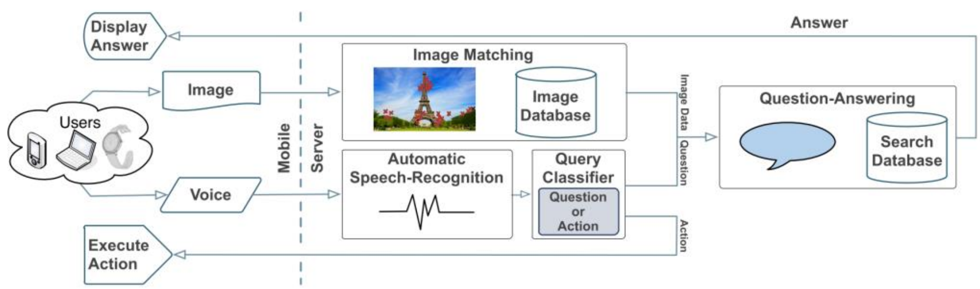


Figure 1 Diagram of The Sirius Pipeline

(Hauswald et al. 2015)

Natural Language Processing (NLP) uses computational techniques to understand and produce human language contents. NLP research focuses on tasks such as speech recognition, information extraction and retrieval, text summarization, question answering, topic discovery, and opinion mining etc. (Cambria and White 2014).

QA system is an essential part of a cognitive assistant. There are three main types of QA systems: IR-based QA, knowledge-based QA and hybrid QA (Gandomi and Haider 2015; Hauswald et al. 2015). IR-based question answering relies on available information such as text on the Web or collected textual data in the knowledge base. Knowledge-based

QA and the hybrid QA are more computational complicated. They apply machine learning models to analyze data sources in the knowledge base, and then select the most relevant information and generate a best answer. IBM Watson's Deep QA system is an example of a hybrid QA, and it has shown that improved QA systems can support professionals to make important and timely decisions in many areas, such as compliance, health care, business intelligence, knowledge discovery, enterprise knowledge management and customer support etc. (Ferrucci et al. 2010; IBM White Paper 2011). Figure 2 (Mehrez et al. 2013) shows the general QA analysis process of a knowledge-based QA system.

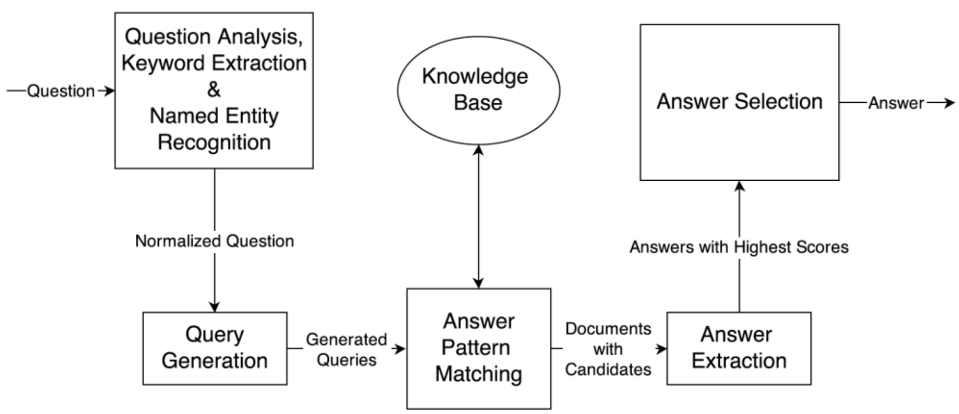


Figure 2 QA Analysis of a Knowledge-Based QA System
(Mehrez et al. 2013)

An adaptive learning module is an advanced function designed in a cognitive assistant to provide recommendations to users (Myers et al. 2007). It trains and adapts itself by interacting with users and learning users' preferences over time. The module is developed in the NLP modules and in the QA System so that the cognitive assistant could continuously learn to better understand users' questions and could predict what a user will need based upon his historical interactions with the system and provide recommendations to him and to other users when similar questions are asked.

1.2 Research Motivation and Research Contributions

At present, audit firms usually use a checklist as the decision support tool for the audit plan brainstorming meetings (Bellovary and Johnstone 2007), which includes a list of risk areas such as entity understanding, industry environment, significant accounts and fraud risks that are required by the audit firm to discuss during the meeting. However, there are some issues with it. Firstly, a checklist may limit auditors' ideas in identifying new potential risks as it is very structurally-restrictive. Secondly, in the brainstorming meeting, conversational contents on how the engagement team identify fraud risks and make audit decisions cannot be collected. Thirdly, a checklist cannot provide information retrieval support or suggestions to auditors to help them in risk identification and assessment. Therefore, there is much room for the improvement of audit tools in this audit process, and we propose that cognitive assistant technology can be applied to solve this issue.

To build a decision support cognitive assistant, one of the most important steps is the development of knowledge base. It is very challenging to develop such a knowledge base because it should contain all risk assessment related information to support the QA system of the cognitive assistant. The data sources include both structured and unstructured data that might be used by the engagement team, and one of the most valuable unstructured data sources that should be stored is auditors' experience and expertise. Audit brainstorming meeting discussions create such knowledge and help interested parties (such as regulators, practitioners, and academics) better understand how more experienced auditors evaluate information and make risk assessment judgements. Moreover, senior auditors' experience and expertise, once collected, can be used as decision aids for future audit plan engagements.

However, there have been no existing researches on intelligent audit knowledge discovery and analysis from audit brainstorm meeting conversations. The current knowledge discovery of meeting contents relies on manual process of meeting summaries, which is inappropriate for future content analysis and related knowledge base development. There are no developed systems or framework in the existing NLP literature that can process audit conversations, and no existing audit tools have been developed on intelligent audit risk assessment knowledge discovery.

Although the information from the brainstorming meeting has great value, it is very challenging to retrieve the knowledge from the audio streaming of the auditors' discussions. Firstly, it is difficult to extract the main contents from a long, unstructured audio streaming. Secondly, it is challenging to define and discover risk assessment related knowledge from the extracted contents that can provide decision support for future audit plan engagements. Thus, we propose an NLP-based audit plan knowledge discovery system (APKDS) for knowledge discovery from audit conversations.

The actual development of such an intelligent audit conversation analysis system is technically challenging as well. The development of NLP models involves machine learning tasks and needs experimental audit brainstorming discussion data to train and test the models. To realize the proposed NLP-based audit conversation analysis system, we propose an APKDS prototype and demonstrate the development of some key modules by integrating NLP and machine learning techniques, and then train and test the modules using experimental data.

This dissertation is an attempt to respond above issues. Generally, the proposed collection of research has the following major contributions:

- This is the first study that applies cognitive assistant and cognitive computing technologies to accounting and auditing domain. A general framework is proposed for developing an intelligent audit cognitive assistant that auditors can interact with and get decision aids during the audit plan brainstorming. It will not replace existing computer-assisted audit techniques (CAATs) but work as a supplement to them. Moreover, the audit cognitive assistant can be extended to support other important audit phases, such as client acceptance, preliminary engagement activities, internal control audit, business processes and accounts audit, and audit completion.
- The proposed audit cognitive assistant provides a new method of audit knowledge organization, which potentially creates a knowledge base that contains many senior auditors' knowledge and experience in audit risk identification and assessment.
- This is the first study that develops a method that can automatically and effectively collect and analyze audit conversational contents. The proposed NLP audit plan conversation analysis framework can extract main brainstorming contents and transfer them into knowledge for future use. The extracted knowledge explains how auditors make decisions during the audit plan, which provides insights to researchers who study on auditor behavior related topics.

1.3 Overview

The remainder of this dissertation is as follows: The three essays are contained in chapters two, three and four. Chapter two focuses on proposing the audit domain cognitive assistant framework that can provide interactive decision support to audit plan risk assessment. Then a demo is created to demonstrate the development of an Information Retrieval-based QA system. Chapter three proposes the NLP audit plan knowledge discovery system for collecting and extracting important brainstorming discussion contents automatically and continuously. Chapter four proposes the development of a prototype for the audit knowledge discovery system. Experimental brainstorming meeting recordings are used to train and test the models. In the end, we discuss how to extend the design of the tool from audit plan brainstorming to other audit stages. The last chapter concludes the dissertation and pointing out future research areas.

Figure 3 shows the relationships of Chapters 2-4.

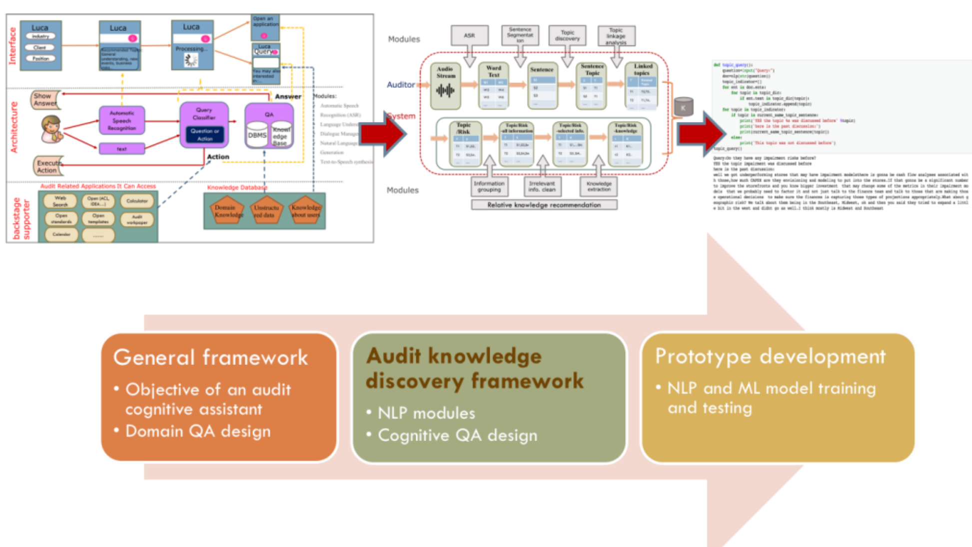


Figure 3 Relationships of Chapters 2-4

CHAPTER 2: DEVELOPING A COGNITIVE ASSISTANT FOR THE AUDIT PLAN BRAINSTORMING SESSION

2.1 Introduction

The intrinsic nature of auditing allows for the use of audit software which can provide auditors assistance in a variety of tasks. In today's information world, auditors need to use many data sources to understand and evaluate the industry and the business of audit clients. In addition to traditional financial information, auditors should expand to non-financial information from external sources such as news articles and social media. Accounting and auditing are changing fundamentally due to the technology improvement in data analytics and Artificial Intelligence (Kokina and Davenport 2017).

Audit firms have nowadays been investing many resources into AI-related projects (Greenman 2017). KPMG signed a broad agreement with IBM to apply IBM Watson to a series of audit processes (Lee 2016). KPMG's cooperation with Watson tries to develop selected cognitive services designed to help KPMG "meet its extensive audit-specific security, confidentiality and compliance requirements" (IBM 2016). Deloitte is trying to assemble different cognitive capabilities from various vendors and integrate them to support audit processes, such as document review and predictive risk analytics (Raphael 2017). PwC and EY are increasing their usage of audit platforms and predictive analytics (Kokina and Davenport 2017).

Cognitive Assistants or Intelligent Personal Assistants (IPA) are speech-enabled technologies that use a natural spoken language and semantic understanding techniques to communicate and interact with human and help human obtain wanted information (Canbek and Mutlu 2016). These tools allow users to input information such as the user's voice, vision (images) and contextual information, and then provide instant assistance to users through answering questions and performing actions (Hauswald et al. 2015). Examples of commercial IPAs include Apple's Siri, Google Now, Microsoft Cortana, Amazon's Echo/Alexa, IBM's Watson, etc.[1] (Canbek and Mutlu 2016; Mehrez 2013; Ebling 2016; Bellegarda 2013; Strayer et al. 2017). Current cognitive assistants have shown their usefulness in people's daily lives, but the technology has been slow to be applied to the work environment. Cognitive assistants are great tools for business situations when users need information retrieval support from a large amount of knowledge sources.

This study proposes that cognitive assistant technology can be applied in the accounting and auditing domain. During audit planning and risk assessment, a critical step is usually to hold brainstorming meetings (Bellovary and Johnstone 2007). During the brainstorming stage, engagement team members exchange ideas about how and where they think the client's financial statements may be susceptible to material misstatement due to fraud or error (Beasley and Jenkins 2003). Considering the importance of the brainstorming session and the limitation of existing audit decision support tools for the process (Dowling and Leech 2007; Seow 2011; Landis et al. 2008), in this paper, a cognitive assistant framework is developed for the audit plan brainstorming session with the objective of providing timely information retrieval and decision-making support for audit engagement team members. In addition, a demo is developed to show one of the main modules (IR-

[1] Other commercial IPAs are BlackBerry's 'BlackBerry Assistant', Braina, HTC's 'Hidi', Maluuba Inc's 'Maluuba', Motorola's Mya (unreleased), Samsung's 'S Voice', Cognitive Code's 'SILVIA', Nuance's 'Vlingo', LG's 'Voice Mate' (Canbek and Mutlu 2016)

based Question Answering module) in the proposed cognitive assistant, and it is tested with data extracted from experimental brainstorming meetings.

This study will contribute to the literature and practice in three main aspects. First, this is the first study discussing how cognitive assistant and cognitive computing technologies can be applied in accounting and auditing domain. Second, this study proposes a general framework for developing an intelligent audit cognitive assistant to support auditors in audit plan. Third, the proposed tool provides a new method of knowledge organization for the audit domain, which potentially develops a knowledge base that stores many senior auditors' knowledge and experience in audit risk identification and assessment.

The structure of this paper is as follows. Section 2 discusses the importance of the audit brainstorming session and its current issues. Section 3 describes the features of cognitive assistant technology and why it can be used to support audit brainstorming. Th main methodology of the study is introduced in section 4. Section 5 introduces the general framework of the proposed system, including the framework and important computer modules of the system. Section 6 describes in detail how to design the QA system and recommendation system. A demo for the Information-Retrieval (IR) based Question Answering module of the proposed system is developed and presented in Section 7. Last section is a summary.

2.2 Background on Audit Brainstorming Sessions

2.2.1 Audit Brainstorming Sessions

Two auditing standards mandate brainstorming sessions, SAS No 99 and SAS No 109 (AICPA 2002). SAS. 99 requires auditors identify specific fraud risks and SAS 109 requires auditors identify additional causes of potential material misstatement in the financial statement during brainstorming (Landis et al. 2008; Carpenter 2007; Hoffman and Zimbelman 2009; Bellovary and Johnstone 2007; Hunton and Gold 2010; Lynch et al. 2009). The participating team may include the whole audit team, including audit partner, audit manager, senior and new staff, and the time spent on the session varies from 15 minutes to 2 hours, mainly 30 minutes to 1 hour (Bellovary and Johnstone 2007).

Studies on brainstorming session generated ideas on how to improve the effectiveness of the meetings. Some studies developed guidelines for better brainstorming, such as establishing ground rules, setting the encouraging tone, and encouraging more ideas not less etc. (Landis et al. 2008; Beasley and Jenkins 2003). Beasley and Jenkins (2003) discussed what format of brainstorming works best. The most common technologies include open brainstorming, round-robin brainstorming and electronic brainstorming, and each format has its own advantages and disadvantages. Ramos (2003) described how to structure an effective brainstorming session and provided basic rules to encourage participation.

Brainstorming is very important step in the audit plan for in several aspects. First, literature showed that auditors' fraud identification and assessment plan efforts affect their subsequent performance of audit, including evaluation of evidence and final fraud risk assessment (Carpenter 2007). Second, brainstorming teams tend to generate more high-quality ideas on risks than individual auditors generate by themselves (Landis et al. 2008; Beasley and Jenkins 2003; Carpenter 2007; Hoffman and Zimbelman 2009). Audit team's risk assessment after the brainstorming session are significantly higher than assessment given by individual auditors before the session (Carpenter 2007). Studies showed that auditors are more likely to identify correct risk considerations if an audit team engages in an open-ended discussion. Also, some senior auditors have more experience and expertise, while some juniors have less experiences but have recent first-hand knowledge of client processes. Effective discussions among experienced auditors and junior team members can encourage junior auditors better share their insights with the team (Beasley and Jenkins 2003). Third, auditors must document the risks identified during the brainstorming session and respond with modifications to the audit plan that address identified risks. Documentation of the risks becomes a decision aid that facilitates subsequent information retrieval (Nelson 2009).

SAS No. 99 didn't specify how auditors should conduct brainstorming meetings or what training and guidance should be available to auditors (Bellovary and Johnstone 2007). Inefficiencies and distractions of brainstorming sessions may ultimately muddy the audit team's ability to identify risks and hinder key audit decisions, leading to dangerous paths (Beasley and Jenkins 2003).

2.2.2 Issues of the current audit brainstorming audit tool

An audit firm usually use a checklist as the decision support tool during audit plan brainstorming sessions (Bellovary and Johnstone 2007), but there are some issues with this decision support tool (Dowling and Leech 2007; Seow 2011; Landis et al. 2008).

Firstly, checklists may limit auditors' ideas in identifying new potential risks. A checklist can be considered as a structurally-restrictive decision aid tool as it shows a list of risk areas to guild brainstorming discussions. Seow (2011) shows that the more structurally-restrictive decision aid imposes more limits on users' decision-making process and induces biases because users are forced to adapt their decision-making to match the guidance in the tool. This type of decision aid reduces users' ability to identify new items that did not appear in the tool when users focus only on the items prompted by the tool and fail to adequately consider other possibilities in a particular situation (Seow 2011; Asare and Wright 2004; Dowling and Leech 2007). A checklist usually provides only common discuss items for the engagement team without specifying industry or client specific risk areas, thus auditors may only focus on the suggested items shown in the checklist and fail to identify other possible risk factors.

Secondly, in the brainstorming meeting, conversational contents on how the engagement team identify risks and make audit decisions cannot be collected. Senior auditors' expertise and experience about industries and clients could be a valuable information source for future similar engagement cases. Taking meeting notes may be the easiest and most commonly used method for tracking a meeting, but the contents written down will be very limited.

Thirdly, a checklist cannot provide information retravel support. Auditors need to recall information from memory or look for information from pertinent documents, which take time and require much preparation work. In addition, if auditors need information about the case from other external databases, they cannot access to it easily. Therefore, although a checklist is an essential and accessible decision support tool for current audit plan process, there is much room for the improvement of audit decision support tools.

2.3 Features of Cognitive Assistants and Their Advantages in Improving Audit Planning

2.3.1 Features

Cognitive Assistant can be a good solution for improved audit plan brainstorming for two main reasons. First, technology improvement makes it feasible to build such kind of tools. As AI technologies become much more mature and applicable, new technologies in cognitive assistant can fit the need of improving brainstorming session effectiveness and efficiency. Since the proposed intelligent assistant must be trained with large training dataset, in recent years, necessary big data support became possible when companies have been generating increasing amount of data records due to improved ERP system. Second, the features of cognitive assistants make it a good solution to a success of an audit brainstorming meeting.

At present, the most important benefits of cognitive assistants include offering information retrieval, supporting users with recommendation systems, adaptive learning capability and service delegation systems (Garrido et al. 2010; Myers et al. 2007). The benefits of cognitive assistants make them excellent tools for the audit field, especially the processes where auditors need information and decision support based on a massive amount of data sources. Figure 4 shows the main features of a typical cognitive assistant and how these features can benefit audit brainstorming sessions.

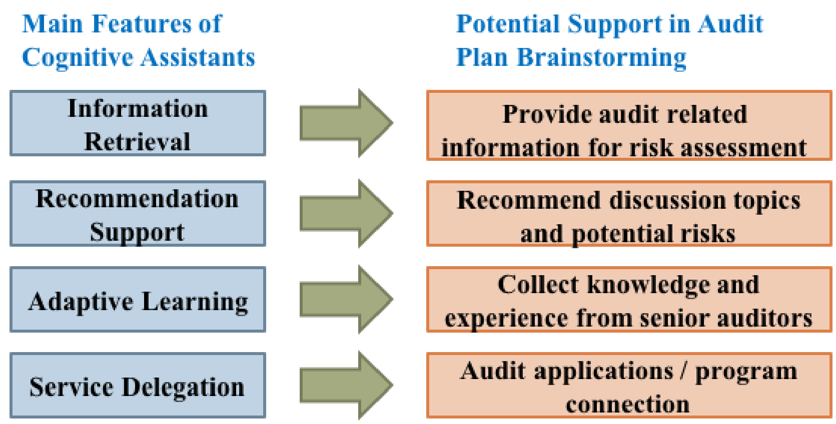


Figure 4 Features of Cognitive Assistants and Benefits to Audit Brainstorming

Firstly, a cognitive assistant can provide information retrieval support to the engagement team at industry and client level. Primary information sources for audit brainstorming such as financial statements, regulation, analytical procedures guidance etc. can be stored in the knowledge base of a cognitive assistant. Auditors should be able to ask questions to the tool to receive answers.

Secondly, a cognitive assistant can provide recommendations to users for decision making support through its recommender system. Since this tool is proposed to provide decision support for the audit plan brainstorming, suggestions on risk areas and discussion topics for a specific engagement case should be provided to auditors. Initial recommendations will be needed to set up a recommender system.

Thirdly, there is a learning curve with a cognitive assistant. The adaptive learning feature makes the system adapt itself by interacting with users and learning users' preferences over a wide range of functions within the system (Myers et al., 2007). For instance, Google Now can recognize repeated actions that users perform such as repeated calendar appointments and search queries, and therefore can display more relevant information to the user. There are two types of “learning” capabilities in the system that can help improve the tool. Firstly, the system will collect what has been “searched” from different senior auditors, and it is supposed to provide improved recommendations on vital risk areas and discussion topics when more auditors use it. When more auditors use the tool, experience and domain knowledge from senior auditors will be collected through their interactions, and the tool will increase its knowledge base and adapt itself with numerous auditors’ “thinking” and become smarter. Secondly, from natural language processing perspective, the adaptive learning function allows the system to improve its language understanding capability through the model training process with various expressions used by different auditors in the question-answering interactions.

Fourthly, Cognitive Assistants have service delegation systems (Garrido et al. 2010; Canbek and Mutlu 2016; Ebling 2016), which can invoke various apps to complete tasks such as setting a timer, sending messages, getting directions, creating reminders, reminding daily appointments, responding to questions (Canbek and Mutlu 2016). A cognitive assistant for audit planning could complete tasks such as calculating, returning data analytical results from related data analytical programs, locating needed information from documentations, searching for external information (from Internet or external databases), creating meeting schedules etc. These functions save auditors time in the brainstorming meetings and allow the engagement team members focus more on important tasks in the risk assessment discussions.

2.3.2 Challenges

There are some challenges of developing the cognitive assistant for auditing and accounting. The first type of challenge comes from general challenges of cognitive assistant development, including:

- many backs and forth in commands;
- users are required to listen intently for long time for answers to be read;

The second challenge is the collection of knowledge from auditor expertise. To build a decision support cognitive assistant, one of the most important steps is the development of knowledge base, which is supposed to be built through interactions between system prototype and experimental users in the training process. Therefore, it will be necessary to have a group of experienced auditors as initial expertise, and their knowledge should be collected and analyzed by the cognitive assistant to start building its knowledge base and question answering system.

The third challenge is the heterogeneous information integration. To be able to provide information retrieval and risk assessment support, the proposed tool needs to integrate different information sources from internal and external databases, and prepare them in a standardized format that can be used for further extraction and analysis.

2.4 Methodology

The audit cognitive assistant framework proposed in this study was developed using design science (Gregor and Hevner 2013; Hevner et al. 2004). Design science is a research methodology which tries to create new knowledge or understanding with the objective of solving real-world problems through designing novel or innovative artifacts (e.g., constructs, processes, algorithms, methods, and frameworks) and evaluating the artifacts (Hevner et al. 2004). Gregor and Hevner (2013) described that there are four main types of knowledge contribution that a design science research project can have: invention (new solutions for new problems), improvement (new solutions for known problems), exaptation (known solutions extended to new problems) and routine design (know solutions for known problems). This study falls to the category of Improvement which a better solution (audit domain cognitive assistant) is created to improve the effectiveness and efficiency of risk assessment in audit plan brainstorming meetings.

The development of the proposed system follows the general design-science research guidelines (Hevner et al. 2004). Based on the current practices in the audit brainstorming meetings process, this study aims to develop a feasible cognitive assistant which works as a decision support tool for audit brainstorming discussion process (Design as an Artifact). The framework is intended to address the issue that the engagement teams are required to effectively identify business risks and fraud risks based on their memories and huge amount structured and un-structured data in very limited time during the brainstorming meetings (Problem Relevance).

The proposed cognitive assistant framework (structure and modules) is developed based on review and analysis of currently available commercial cognitive assistants and literature on cognitive computing and audit brainstorming meetings. To build the knowledge base for the question answering system module in the cognitive assistant, information from experimental brainstorming meetings was extracted and used (Design as a Search Process). The framework should then be evaluated and modified through the prototype method (Design Evaluation and Communication of Research) (Hevner et al. 2004).

This cognitive assistant is supposed to be developed with emerging technology which provides audit firms customized risk assessment audit plan decision support for clients of different industries, and it is realizable under current technologies (Research Contributions and Research Rigor). Once the initial version of the system is developed, the system knowledge base will expand continuously through user interactions due to the nature of cognitive computing, which is also an improving and evaluating process.

2.5 Proposed audit cognitive assistant framework

Figure 5 is the proposed architecture of the audit domain cognitive assistant. The proposed audit cognitive assistant framework includes three main parts: user interface, cognitive assistant architecture and knowledge base.

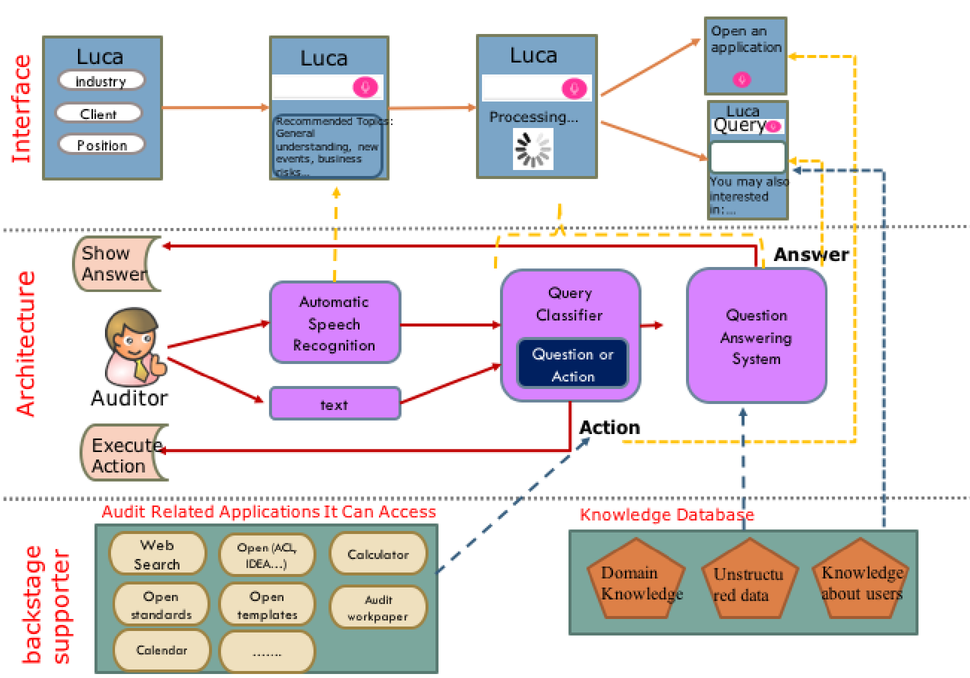


Figure 5 Proposed Framework of the Audit Cognitive Assistant

2.5.1 User Interface

The natural language user interface allows system users to interact with the system. The proposed tool is named as Luca, and users should be able to start the tool when calling Luca. When a user starts a brainstorming meeting, he/she is supposed to select the industry, client company name, and his position in the engagement team from the home page.

Industry and company name should be selected at the beginning for two main reasons: access control and recommendation support. First, in an audit engagement case, the engagement team members should only be authorized to access to information sources that relevant to the case, while data about other clients stored in the knowledge base should not be accessible by these auditors. The selection of industry and client name is the control for information access. Second, since data sources in the knowledge base are supposed to be categorized and tagged, the recommender system and question answering system will process faster in preparing answers to selected industry and client.

By choosing the position of the user, such as partner, manager, junior auditor, tax expert, IT expert etc., the cognitive assistant will be able to learn and memorize queries and actions performed by different auditors, then provide relevant and customized recommendations to following users. For example, if the audit manager is new to the team, when interacting with the tool, he/her will receive recommended discussion topics that raised by other senior auditors when they assess risk for this client or similar clients in this industry in the prior audit.

Examples of the question- answering interaction can be:

Q: Where are the new stores that xxx opened this year?

A: There are five new stores for xxx in 2016:

Three stores in Mexico: xxx

Two stores in Canada: xxx

Q: List the new acquisitions and new joint ventures of xxx this year.

A: There is one acquisition in xxx in 2016: xxx

Q: Does xxx have new suppliers?

A: There is one new partner in xxx: xxx

Q: Does xxx have new contracts or covenants?

A: There are no new contracts or covenants found in 2016.

2.5.2 Recommender System

Since this audit domain cognitive assistant is proposed to provide decision support for the audit engagement team during audit plan and risk assessment, it is crucial that it can give domain knowledge for decision support. The recommendation system's knowledge base should be built based on audit standards, procedures of the audit firm, and knowledge that the cognitive assistant learned from user interactions (through auditors' queries). Therefore, as more auditors interact with the tool, it will react better by providing more targeted recommendations. For example, the recommender system should include many pre-programmed risk areas that an engagement team usually discuss during a brainstorming meeting. Based on selected industry and client, the system should select some important risk areas from all the risks using its ranking model and return those to the engagement team as recommendations.

For each risk area, the system can be designed with more detailed discussion topics to recommend to auditors. For instance, suggested discussion topics under General Understanding of the Firm can be business operations, management, investments, etc. Since each industry may have industry-specific risk areas and discussion topics to focus, both general topics and industry-specific topics should be designed in the recommender system.

2.5.3 Architecture

The middle layer is the proposed architecture of the cognitive assistant. Auditors can interact with the proposed tool by talking to it or typing in the question that they want to ask. Automatic Speech Recognition (ASR) will process auditors' voice and then translate his/her voice question into its text equivalent through analytical models. The translated text then goes to Query Classifier which decides if the command is an action or a question. If it is a question, then command goes to the Question Answering (QA) system. QA system will extract information from the question, search its database, and choose or generate the best answer and return to the user. For example, when auditors discuss how weather and seasonality issue may affect the sales performance of stores of a client, they may want to ask the cognitive assistant "which stores may be affected by weather?". Then the tool should look for weather-related information from historical records and prepare an answer. The answer could be a list of stores locations that were affected by hurricanes (such as south, southeast, east coast) and those not in prior years, together with related insurance coverages and losses. Also, the answer may also include predicted risky store locations based on recent weather news and predictions. If it is an action, the command is sent back to the system to execute the required application.

This architecture is proposed based on the structure of Sirius (Hauswald et al. 2015). Sirius is an open intelligent personal assistant system built by University of Michigan engineering researchers with both speech and image front-ends. Sirius works similar to Siri, Microsoft Cortana and Google Now, but besides accepting voice commands, it allows users to ask questions about what they are seeing (Hauswald et al. 2015). However, considering the information needs of auditors during audit plan brainstorming, image recognition function in Sirius is not considered in the proposed tool.

2.5.4 Applications

Applications are the third-party apps that are linked to the proposed audit cognitive assistant for direct execution. During the audit brainstorming meetings, auditors use information from various documents to discuss risks and make judgments. Therefore the service delegation function in the cognitive assistant becomes an "information and task manager" that helps improve the work efficiency of auditors during the process and make them focus more on important tasks such as risk discussions. For example, auditors may be interested to know new information such as economic factors and latest events about a client during the risk assessment, and an app of "web search" allows information that cannot be found in the current knowledge base to be obtained through existing online search engines. Only audit plan related applications should be linked to this proposed tool. Capabilities of some proposed relevant apps are listed in the table below (Table 1).

App	Capability

Audit analytical tools (such as ACL, IDEA)	Support audit data analytics such as statistical and predictive analysis
Audit working papers	Provide evidence from prior audit working papers
Calculator	Support financial ratios analysis for financial statements; other data analytical calculations such as Benford's law (Nigrini 2012; Dai and Li 2016)
Templates	Guidance or programs designed by the audit firm for required procedures in risk assessment
Standards	Regulations related to the audit area or financial account
Web search	Direct search from search engine such as Google; used when information cannot be found in existing knowledge bases

Table 1 Proposed Apps

2.5.5 Knowledge base

The knowledge database is used to store audit brainstorming related data sources to support the QA system. The knowledge base should include both structured data and unstructured data that could be used by the engagement team during audit plan risk assessment. Organized knowledge storage free auditors from memorizing tremendous amount of information. It provides efficient information retrieval during brainstorming sessions for question-answering interactions.

Text resources are unstructured data sources stored in the knowledge base. Phrases, sentences, and paragraphs in all text resources will be tagged based on pre-determined question category and answer type to enable the information in those texts to be extracted by algorithms during user query. Data sources include financial statements, accounting policies, analytical procedures, litigation, claims, recent news information, audits working papers, prior year audit deficiencies and adjustments, etc. For example, the amount of a firm's Accounts Receivable - Net in the Balance Sheet should be tagged based on several keywords: firm’s name, fiscal year, annual financial report, balance sheet, and accounts receivable - net. Then when an auditor generates a question such as “what is xxx (firm’s name)’s accounts receivable in the annual financial statement of 2016?”, the tool will know that the question category is of this query is “Annual Financial Report”, and then look for best answer from the corresponding sub-knowledge base.

Domain knowledge means collected experience and expertise from auditors. Some of these insights are essential facts about client financial situation that could be collected and prepared before a discussion meeting. Some knowledge contains judgments or experience that were extracted from prior audit documents, which provide important insights for new cases. This domain knowledge can be prepared and stored in the cognitive assistant as questions and answers, and they should be stored as structured data a relational database.

The third type of knowledge is new knowledge gained through user interactions with the tool. Based on user’s queries and search behaviors, related new knowledge about users is collected for future recommendations in similar situations.

2.5.6 Important modules in the proposed system

To develop a cognitive assistant for the audit domain, one strategy is to utilize from public computer domain modules. The audit plan cognitive assistant framework in this paper is proposed based on architecture and modules of current commercial cognitive assistants. The most important components used for speech-driven cognitive assistant include Speech Recognition, Spoken Language Understanding, Language Generation, Question Answering System, Dialog Manager (Query Classifier), and Text to Speech

(Mehrez et al. 2013; Bellegarda, 2013; Britta, 2015; Hauswald et al. 2015). Figure 6 shows an example of how a current cognitive assistant is structured (Mehrez et al. 2013). In their architecture, language understanding and language generation are implemented within its question-answering system. The dialog manager module is used to decide if the user's command is an action or a question. The text-to-speech component will transform answers in natural language text to voice responses.

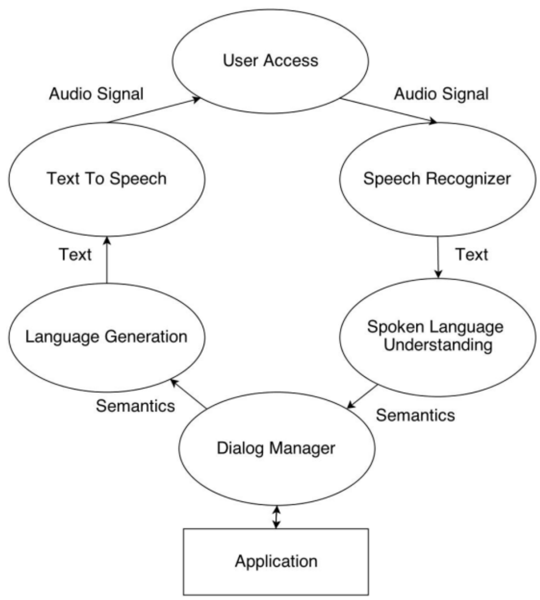


Figure 6 A Basic Cognitive Assistant Architecture

(Mehrez et al. 2013)

a. Speech Recognition

Intelligent personal assistants such as Apple's Siri and Amazon's Alexa leverage machine learning techniques, including Deep Neural Networks (DNN), convolutional neural networks (Tang and Lin 2017), long short-term memory units (Chen et al. 2015), gated recurrent units (Ravuri and Stolcke 2016), and n-grams (Daelemans 2013; Levy, 2016) to build smart voice recognition system. These techniques can be used in building the cognitive assistant system for audit domain as well. Developers can also integrate components from well-established open source projects which include methods and algorithms used in commercial systems. For speech recognition, popular open projects include Carnegie Mellon University's Sphinx (Gaussian Mixture Model based) (Huggins-Daines et al. 2006), Microsoft Research's Kaldi (Povey et al. 2011) and Germany's RWTH Aachen "RASR" (Deep Neural Network based) (Rybach et al. 2011).

b. Language Understanding and Language Generation (Natural Language Processing)

Since the 1950s, Natural Language Processing (NLP) research has been focusing on tasks such as machine translation, information retrieval, text summarization, question answering, information extraction, topic modeling, and opinion mining (Cambria and White, 2014). It uses computational techniques to learn, understand, and produce human language content. Speech recognition and speech synthesis are part of NLP, and it is widely used in real-world applications, including creating spoken dialogue systems and speech-to-speech translation engines (Hirschberg and Manning 2015).

With the help of NLP, voice commands and natural language questions from auditors should be translated into “parsed text” so that the computer programs can understand. In the proposed cognitive assistant, NLP should be implemented within its question answering system so that it can understand auditors’ questions and generate answers.

c. Question Answering (QA) system

Question answering (QA) techniques provide answers to questions that rely on complex NLP techniques. QA systems have been implemented in healthcare, finance, marketing, and education. Apple’s Siri and IBM’s Watson are examples of commercial

QA systems. QA system is one of the most critical functions of the proposed intelligent system. QA system development has a long history, with three main modern paradigms of question-answer systems: (1) IR-based question answering; (2) knowledge-based question answering; and (3) hybrid approach question answering (Gandomi and Haider, 2015). Apple's Siri is an example of a knowledge-based approach. In hybrid QA systems, like IBM's Watson, while the question is semantically analyzed, candidate answers are generated using the IR methods (Gandomi and Haider, 2015). One application of Watson is in the medical industry where it analyzes patients' medical information against a considerable amount of data and expertise to offer evidence-based treatment options (Ebling 2016).

One crucial feature of Watson's Deep QA is its cognitive computing technology. Cognitive computing refers to systems that learn at scale, reason with purpose, and naturally interact with humans. It can learn, reason and improve machine "knowledge" from their interactions with humans (Goffman, 2016). Cognitive technology allows greater collaboration between humans and systems, providing the ability to communicate in the natural language and analyze massive amounts of data to deliver insights more quickly. There had been a trend of applying this technology into accounting and auditing domain. Many audit, tax, advisory and other services rely heavily on judgment-driven processes. Adding cognitive technology's massive data analysis and innovative learning capabilities to these activities has the potential to advance traditional views on how talent, time, capital and other resources are deployed by professional services organizations (IBM 2016).

To satisfy the needs of question answering in the audit plan risk assessment in brainstorming meetings, the question answering system in the proposed cognitive assistant should involve both knowledge-based QA (cognitive QA) and IR-based QA (domain QA). The proposed QA system architecture (Figure 7) is built based on existing QA models (Mehrez et al. 2013; Ferrucci et al. 2010; Chung et al. 2004).

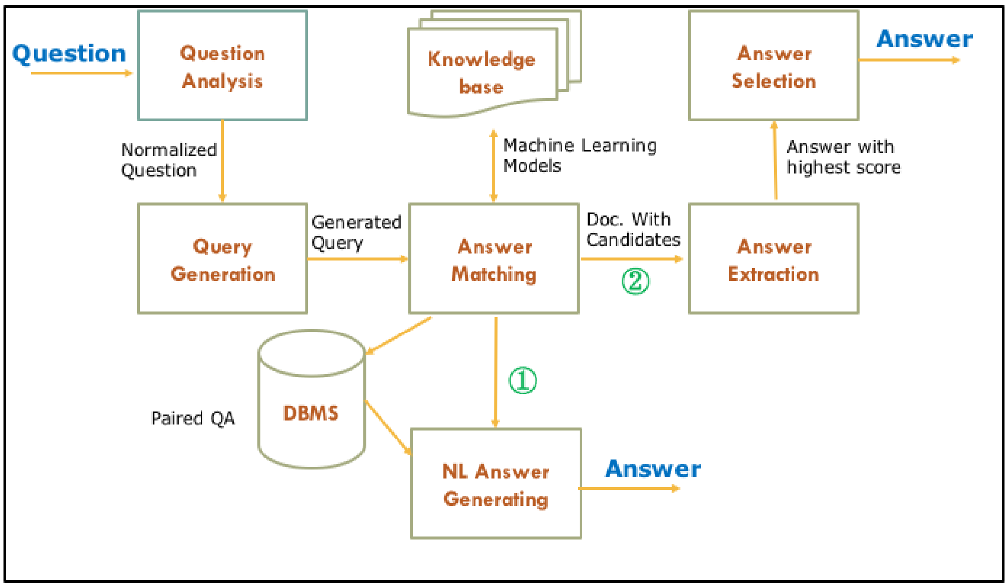


Figure 7 Proposed QA System

In the proposed architecture, first of all, the questions are normalized to be accessible for query generation. Normalization includes question analysis, keyword extraction and named entity recognition to generate a proper query for this question. Punctuations are removed, abbreviations are expanded and nouns and verbs are stemmed. Then the generated query will be used to search from the knowledge base. With the extracted type of question, Answer Matching method is used to find possible answers. If the QA system can find pre-paired answers stored in a relational database like traditional domain QA systems (Chung et al. 2004), the NL answer generating module will generate an answer to the user (path 1 in Figure 7). If there is no pre-prepared question type or answers for the question, cognitive QA module will be used to analyze relevant information from the knowledge base, and then use machine learning models to generate documents with answer candidates. Then, candidate answers are extracted and ranked, and answer with the highest score will be returned to the user (path 2 in Figure 7).

Open-source components from heterogeneous sources can be utilized in the development of the proposed question-answering system. For example, the OpenEphyra (Seide et al. 2011) framework is an open-source framework representing the state-of-the-art QA system based on IBM's Watson (Ferrucci et al. 2010). The NLP techniques of OpenEphyra are used at Google and in QA systems of other industries (Täckström et al. 2013; Hauswald et al. 2015).

2.6 Design of QA system and recommendation system

2.6.1 QA system design - question categories

To build the QA knowledge base, first we need to know what types of queries should be prepared with paired (information-retravel based) QA, and what types of questions cannot be answered directly and cognitive QA will be used to analyze from knowledge base. Table 2 shows highest level of question categories that we could build in the IR-based QA system. These question categories are extracted and summarized based on related literature (Messier et al. 2016) and Audit Standard 2100 Audit Planning and Risk Assessment (PCAOB, 2010). The proposed categories only provide a starting point for QA system development, and the question categories and sub-categories can be extended based on needs. For example, Huang and Li (2011) developed a text classification algorithm that classified risk factors in section 1A of 10-K form into 25 risk types (e.g., international risks, regulation changes, shareholder's interest risks etc.), and audit firms may want to add some risk types to supplement the system later. Table 2 shows the initial design of the 13 highest level question categories for QA system knowledge organization. These proposed categories are most important topics and risk areas for the audit plan brainstorming discussions.

Category No.	Question categories (risk areas)
1	Entity Understanding
2	Industry Condition
3	Regulatory Environment
4	Business Objectives and Strategies
5	Significant Accounts
6	Controls
7	Fraud Risks
9	Specialists/ third-party
10	Materiality
11	Related Parties
13	Going Concern
	……

Table 2 Highest Level Question Categories

For each topic above, there are secondary question categories that can be designed to improve audit knowledge collection and classification in the QA system. Table 3 shows some example sub-categories.

Level-1 categories	level-2 categories
Entity understanding	Background, Business operations, Management, Investments, Financing activities, Financial reporting, Multi-locations, External factors,

	Entity performance measures
Industry condition	Competitive environment, Product technology, Cyclical or seasonal activity, Energy supply cost
Regulatory environment	Accounting principles, Applicable financial reporting framework, Legislation and regulation, Taxation (corporate and other), Government policies, Environmental requirements, Violations of laws and regulations
Business objectives and strategies	Market, Reputation, Industry developments, New products and services, IT security, Expansion of the business, Current and prospective financing arrangements, Management's strategies
Significant accounts	Revenue, cash flow, account receivables, liabilities/loans, rent , inventory, fixed assets/intangibles, investment, expenses, payroll, tax etc.
Fraud risk	Fraudulent financial reporting, Misappropriation of assets
Specialists	Specialists in finance, Specialists in tax, Specialists in valuation, Specialists in pension, Specialists in IT
Related parties	Nature of the relationships, Transactions

Table 3 Level-2 Categories

Based on the audit standard about risk assessment and related literature, level-3 question categories that should be designed from some of the level-2 categories above to further classify risk assessment information. Table 4 shows some example level-3 categories.

Level-2 categories	Level-3 categories

Entity performance measures	Financial key indicators, nonfinancial key indicators, budgets, variance analysis, performance reports, comparisons of performance
External factors	General economic conditions, interest rates, inflation
Violations of laws and regulations	Illegal acts, violations of the securities acts, environmental protection, equal employment regulations, antitrust violations
Revenue	Sources, amounts, manual or automated collection/adjustment, new customer contracts, new suppliers, changes in standards, international operation & sales (centralized vs decentralized), control effectiveness (any tools), seasonal routine revenue pattern, materiality, specialist involvement, fraud risks, tests

Table 4 Level-3 Categories

2.6.2 Recommender system

The proposed recommendation system should be built based on three types of information:

- Pre-defined recommendation: sub-question categories;
- Industry specific recommendations (e.g., special expertise of the engagement team for a specialized industry);
- Updated related topics discovered from adaptive learning module

Before the recommender system can learn and update related topics through user interactions, to start, we need a pre-defined recommendation list to create the initial recommended topics for the recommender system.

Since each industry may have industry specific discussion topics, and both general topics and industry specific topics should be recommended to auditors. Many of these pre-defined recommendations can be selected directly from level-2 and level-3 sub-question categories, as these sub-topics are related risk areas of its higher level (level-1) topics. Some topics may be industry specific, so that they will be suggested to auditors in restricted engagement cases. For example, for Significant Accounts, auditors may want to discuss accounts of impairment, pension, prepaid insurance and deferred revenue for restaurant industry cases. For cases from manufacturing industry, they may want to discuss accounts of earnings per share (EPS), contract management, allowance, restructuring, funds, restricted stock warrant and compensation plan, even though these accounts were not set up as initial general recommendations.

Recommendations should be presented to auditors as part of the query answers. When an auditor performs a query to the cognitive assistant, the recommended topics identified in the recommender system will be returned to the auditor. For example, when the engagement team is looking for information on business operations, topics of management, investments, financing activities and financial reporting etc. will be recommended to auditors as related risk areas. The proposed topics in the above tables can be enhanced and extended when the system is trained by many users through interactions. Audit firms who develop and use this system can tailor these procedures based on their customized needs as well.

2.6.3 Knowledge sources

In order to better organize question categories and identify answers for the QA list of the IR-based QA system, various information sources will be needed. Table 5 below gives examples of what information sources that could be used by audit firms during QA knowledge preparation.

Knowledge sources	Examples
Internal sources (existing files or databases)	- Audit documentation (working papers） • Financial statements such as 10K, 8K filling • Auditor’s report • Audit plan and audit programs • Working trial balance • Adjusting and reclassification journal entries • Audit memoranda (discussions like internal controls, inventory observation, errors identified etc.) - Questionnaire of documenting the entity understanding - Audit standards (e.g., AU-C 315, AU Section 316, Auditing Standard No. 12 etc.); - Files from outside parties (e.g., confirmation of the entity’s bank balance; contracts, lease agreement) - Data extracted from ERP systems (e.g., copies of sales invoices) - Audit programs, CAATTs, checklist and templates - Related prior studies (e.g., risk types identified by Huang and Li (2011) and Bao and Datta (2014))
External sources (existing files or databases)	-Press release in clients’ website; -Social media; -Internet news articles; -Financial analytical website such as Yahoo finance, Statista, Bloomberg
Information needs to be collected	-Important committee meetings -Experts’ experience and knowledge

Table 5 Potential Information Sources for QA System Knowledge Base

2.7 Demonstration: IR-based QA system

2.7.1 Audit knowledge database description

The risk assessment and planning discussion stage comes after gathering all relevant information and obtaining an understanding of the client and its environment. During this discussion, the engagement team will evaluate risk factors, identify significant risks of material misstatement, and discuss the possibility of potential fraud. As a part of this discussion the engagement team members need to recall information from memory, review files about the client, and then analyze the information and identify risks.

In order to propose and design an effective interactive audit cognitive assistant for the brainstorming meeting, it is important to know what audit partner and manger actually do during the brainstorming meetings, including the resources and documents they use, the procedures, the topics discussed, how they identify inherent risks, control risks, and fraud risks based on information on hand, and how they make subsequent judgments and decision on what to test and resource allocation. Therefore, in this study, four experimental audit brainstorming meetings among engagement team members during an audit planning session were conducted to help explain the actual discussions during this process[2].

The experiment results are the recordings of four typical brainstorming discussions conducted in a big audit firm. In each audit brainstorming meeting, an incoming manager and a recurring partner were asked to complete a planning phase risk assessment for a client. The four cases were independent but all were designed close to real meeting environment.

[2] This database is used in the working paper: The Use of Verbal Protocol Analysis to Describe the Planning Risk Assessment Discussion of Audit Partners and Managers, by Helen L. Brown-Liburd, Theodore Mock, Andrea M. Rozario and Miklos A. Vasarhelyi

Four different companies in different industries were discussed in the four brainstorming meetings (see Table 6 below). The discussions are recorded as voice and then converted into texts using existing Automatic Speech Recognition (ASR) tools. Verbal protocol analysis is a “think aloud” methodology designed to investigate complex processes, including audit planning (Ericsson and Simon 1984; Klersey and Mock 1989). It is used in this experiment to capture how evidence is evaluated and how audit risk assessment judgements are made.

Although the four cases were for different companies and the topics and risks discussed in each case were different, the cases shared common procedures and important topics to discuss in the audit plan brainstorming. Table 6 shows brief introduction of the four experimental cases.

No.	Participants	Client Industry	Time to Complete	Notes
1	Partner and Manager	Electronics	120 minutes	•Utilized the firms audit planning brainstorming checklist •Information: publicly available sources (e.g., 10K) and partner’s knowledge of the company
2	Partner and Manager	Retail-Home	60 minutes	•Utilized the firms audit planning brainstorming checklist •Information: publicly available sources (e.g., 10K) and partner’s knowledge of the company
3	Partner and Manager	Equipment rental	60 minutes	•Utilized the checklist, but the partner leads the discussion •Information: workpapers and partner’s knowledge of the company

4	Partner and Manager	Restaurant	60 minutes	•The manager referred to the checklist, but the discussion was more based on the partner's response to manager's questions. •Information: publicly available sources (e.g., 10K) and partner's knowledge of the company

Table 6 Brief Introduction of The Four Cases

2.7.2 Module Testing

This section demonstrates the proposed IR-based QA module of the cognitive assistant by using developed QA pairs. A cognitive assistant demo is developed through Microsoft cognitive services LUIS[3], Microsoft Azure Bot Service[4] and QnA Maker[5]. The Azure Bot Service and LUIS together create the conversational interfaces for audit plan brainstorming scenario, and QnA Maker stores the QA knowledge base for the IR-based QA module. In this demo, the expected input is a natural language question asked by a user, and the output is a best matched answer/response identified from the its QA knowledge base.

QnA Maker is used to build the IR-based audit QA base. It is a cognitive service which allows the creation of knowledge bases from semi-structured content like FAQ (Frequently Asked Questions) documents or pages, product manuals, structured

[3] https://www.luis.ai/home

[4] https://azure.microsoft.com/en-us/services/bot-service/

[5] https://www.qnamaker.ai/

documents, or manually added QAs. The QA model should be trained to interact with users in a natural, conversational way. In this demo, sample QA pairs were created to train the model. These sample data were created based on the discussions of the four experimental cases. Examples of QA pairs used in the demo development are like Table 7 below:

Question	Response
What are the economic considerations	Ups and downs of the industry and market; clientele demographics; gas prices (very sensitive to gas prices: when the gasoline prices spike, sales take a deep)
What IT controls do they have	New head of IT; control over POS systems in the stores; control for secure data transmission from store POS system to servers at IT center
Is there any third-party work for them	They have a valuation research company for the market valuation of the market-based award, and for goodwill and intangible impairment models

Table 7 Examples of QA Pairs

A well-trained model will be flexible to user queries, which allows users to ask with different phrases and expressions, and the system will look for the most relevant question based on its training process and return the best answer.

Azure Bot Service is the platform for building conversation interface of the demonstration. Microsoft Bot Framework is used to build, connect, and manage various developed bots, database and apps, and it allows a developed app to be integrated across multiple channels such as a website, app, Cortana, Skype, Kik and Facebook Messenger etc.

LUIS is a machine learning-based service to build natural language into a solution. It can be integrated with the Azure Bot Service to create an intelligent conversation bot.

Developers can use LUIS to add customizable pre-built apps into the cognitive assistant, such as Calendar, Music, and Devices. For the proposed audit cognitive assistant, LUIS service provides the technical foundation for connecting to other audit apps and tools, which could support the design of the system's service delegation function. In addition, LUIS has adaptive learning capability. Its active learning is used to continuously improve the quality of the natural language processing models. Moreover, its speech recognition service can be added to the proposed cognitive assistant to support voice commands.

The following Figure 8 shows the user interface of the demo (screenshot of the test version through Azure portal[6]) which allows auditors to enter questions. Figure 9 shows user interaction examples through the demo.

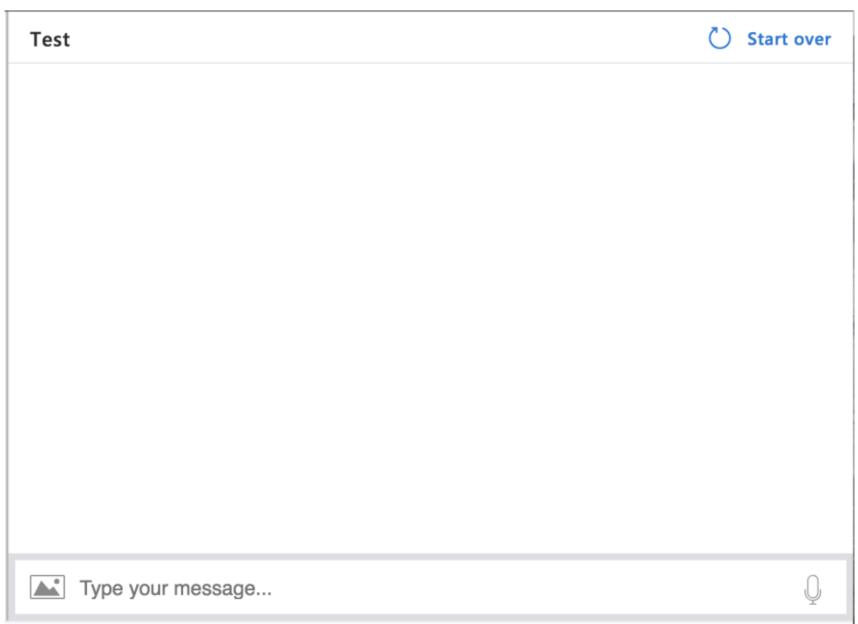


Figure 8 User Interface of Demo

[6] https://portal.azure.com/

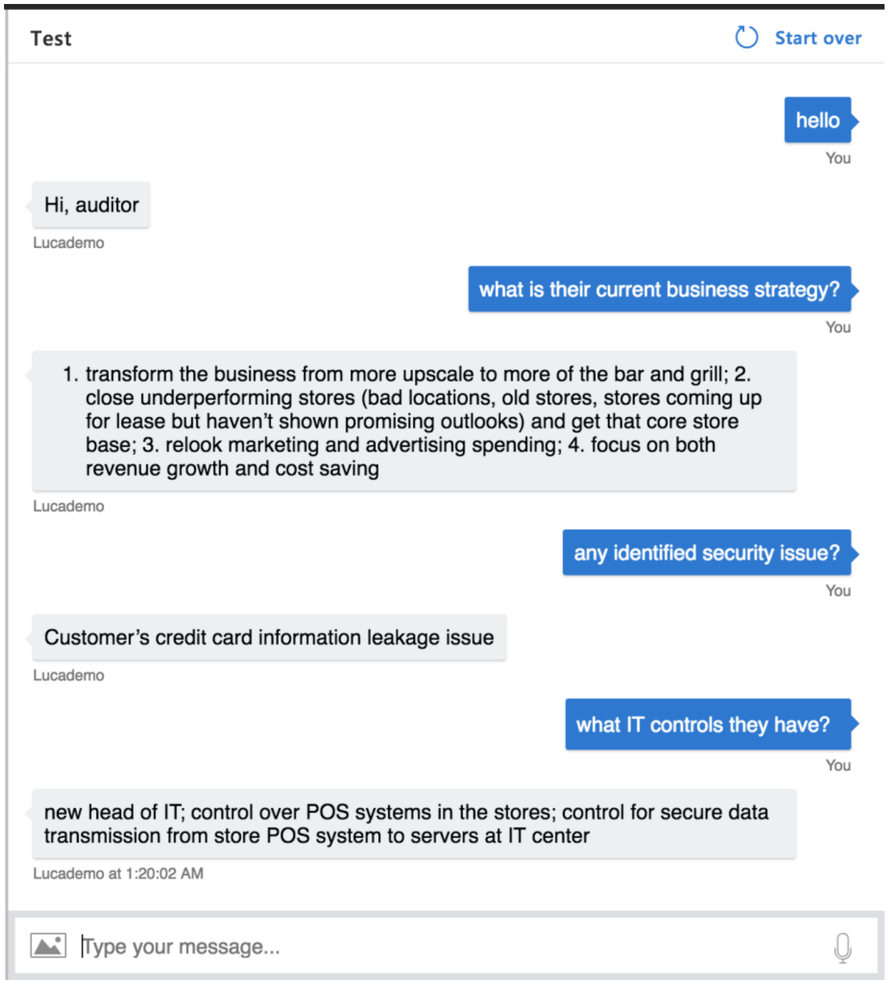


Figure 9 Interaction Examples of Demo

2.7.3 Evaluation Design

Evaluation is a critical part in Design Science Research (DSR) as it assures usefulness and rigor of the artifact (if done correctly) and provides feedback and suggestion for further development (Venable et al. 2016; Prat et al. 2015). Common evaluation criteria in the IS literature are efficacy, usefulness, technical feasibility, accuracy, performance, ease of use, robustness, scalability etc. (Siau and Rossi 2011; Prat et al. 2015; National Research Council 2007; Venable et al. 2016). Peffers et al. (2012) classified artifacts as constructs, models, frameworks, methods, algorithms and instantiations. Based on their summary, common evaluation techniques/methods used by prior literature for a framework include logical argument, expert evaluation, prototype, case study and illustrative scenario.

In our case, to evaluate the proposed audit plan cognitive assistant framework, the most appropriate evaluation method is prototype. The effectiveness of the framework can be demonstrated by a prototype, which serves as the basis for real-world development and evaluation. The demonstrations described above (built with Microsoft language processing services) showed how the proposed cognitive assistant could interact with auditors through a system interface. However, it only provides illustrations of the potential development of IR-based QA system, and it is not a complete prototype which can evaluate all the proposed functions. Peffers et al. (2007) claimed differences of demonstration and evaluation. Demonstration is like a light-weight evaluation which shows that the artifact can work to solve one or more problem instances. It is followed by a more formal and extensive evaluation, which should evaluate "how well the artifact supports a solution to the problem" (Peffers et al. 2012; Venable et al. 2012).

To develop a working prototype for the proposed cognitive assistant, audit firms could utilize popular Natural Language Understanding (NLU) services such as the Microsoft LUIS for its conversational interface. The most popular NLU services are LUIS[7],

[7] https://www.luis.ai/home

Watson Assistant[8], Google Dialogflow[9], wit.ai[10], Amazon Lex[11], and RASA[12] (Braun et al. 2017).

In addition, when a working prototype is built, it needs to be further evaluated to assure the rigor of the design before a complete audit knowledge discovery system is developed with significant investments. Usability valuation provides feedback on whether a design meets user needs. Since initial designs usually rarely fully meet user requirements, proper evaluation process can help avoid the high risk of expensive rework to adapt a system to real user needs or of potential rejection of the system (National Research Council 2007).

Auditors use their professional expertise and experience to identify and assess risks in the audit plan brainstorming sessions, therefore they as users, should be involved in to the prototype evaluation and system development process so that they can provide feedback to system developers on module functions and knowledge base design from user's perspective and work practices. When an early working prototype of the audit cognitive assistant is developed, the most important criteria include effectiveness, accuracy, and satisfaction (National Research Council 2007).

[8] https://www.ibm.com/watson/services/conversation/

[9] https://dialogflow.com/

[10] https://wit.ai/

[11] https://aws.amazon.com/lex/

[12] https://rasa.com/

To evaluate an audit cognitive assistant prototype, laboratory experiment could be a proper method (Hevner et al. 2008; Prat et al. 2015; Peffers et al. 2012; Siau and Rossi 2011). Auditors performance and satisfaction will be evaluated when using the system in a simulated environment. A laboratory experiment is conducted in a well-controlled environment, the researcher decides the place of taking the experiment, time, participants, circumstances and standardized procedure, and the participants should be randomly allocated to independent variable groups. (Siau and Rossi 2011).

To conduct the laboratory experiment, auditors from CPA firms[13] will be the best potential participants. In the experiment, one engagement group could be asked to use the cognitive assistant prototype while another group uses a traditional checklist, and the evaluation could be designed based on their time to finish a brainstorming meeting and the quantity and quality of identified risks. Since the risk assessment and audit decision-making process are subjective, the quality of the risk assessment in the experimental brainstorming can only be determined by audit experts who are familiar with the experiment case scenario based on documented meeting notes after the experiment.

Additional experiment can be designed to understand how the participants think of the usefulness of the new audit tool, compared to the tradition audit checklist. A questionnaire should be used to allow auditors to enter user satisfaction scores on ease of use, functionality, performance, reliability, fitness with the organization, and other relevant quality attributes (Hevner et al. 2008). In general, it is expected that the audit cognitive assistant method will get a higher user satisfaction score than checklist method. Participants

[13] If auditors from CPA firms are not available to be involved in the experiment, capable undergraduate or graduate students in accounting major can be participants

will think that the idea of applying the audit cognitive assistant method into audit plan is promising, and the developed prototype is effective and worth continue developing.

2.8 Summary

AI based cognitive assistants have become increasingly popular as computer-aided tools. This study proposes an audit domain cognitive assistant that can be used in the audit brainstorming meeting to help auditors evaluate information and make subsequent judgments.

Computer-assisted audit techniques (CAATs) or computer assisted audit tools and techniques (CAATTs) have been used by auditors as part of their audit procedures to improve audit effectiveness and efficiency (Janvrin et al. 2008; Mahzan and Lymer 2014). Modern CAATs includes basic office software such as spreadsheet (e.g. Excel) and databases (e.g. Access), general data analytics tools (e.g. SAS), and general audit software (GAS) (e.g. ACL and IDEA).

Different from existing CAATs, the proposed audit plan cognitive assistant provides a new method to manage audit knowledge and a new direction of developing CAATs. This paper is the first study that discusses the potential of applying AI based cognitive assistant technology to auditing. The proposed tool could assist auditors in making better judgments in the audit plan risk assessment by retrieving information and providing recommendations. It is a new type of CAATs and auditors can use it with their exiting CAATs. After interacting with this proposed cognitive assistant during the audit plan risk assessment, an engagement team should obtain better understanding of the client and the risk areas, and then they can use other data analytical tools for further data analysis and testing. Thus, the proposed tool will not replace existing CAATs but work as a supplement to them.

Auditor acceptance of CAATs is an important issue to consider during CAATs design and application (Janvrin et al. 2008). Therefore, in the development of the proposed audit cognitive assistant, developers need to carefully consider the actual information needs of audit engagement teams during audit plan, the functions involved in the tool, and the ease of use.

CHAPTER 3: NATURAL LANGUAGE PROCESSING: AN APPLICATION TO AUDIT KNOWLEDGE DISCOVERY AND ANALYSIS FROM CONVERSATION

3.1 Introduction

Auditing is the process of evaluating and ensuring the truthfulness and fairness of financial statements. High level of knowledge and expertise are required in auditing, which makes it a knowledge-intensive professional service. Many studies stated that knowledge is a critical key to audit success, especially for audit decision making process (Brown-Liburd et al. 2015; Nguyen and Kohda 2017; ICAS 2012). Auditing firms have the need to keep their knowledge management systems updated. Auditors need not only a large amount of knowledge, but also different types of knowledge for high quality decision-making (Bonner 2008). Better judgements can be obtained if auditors know more about how related knowledge is created (Abreu et al. 2014; Bonner 2008; Vera-Muñoz et al. 2006), such as how audit judgements and decisions are made. Brown-Liburd et al. (2015) discussed issues and considerations with audit knowledge in the big data era, which include information overload, information relevance and information ambiguity. There have not been many empirical studies which explained how knowledge is created in auditing (Bouthillier and Shearer 2002; Nguyen et al. 2015; Nguyen and Kohda 2017; Vera-Muñoz et al. 2006). Therefore, a methodology that can help audit firms know better on how audit decisions are made and how risks are identified and assessed can provide great value.

Audit brainstorming meeting in the audit plan is an important process when auditors create knowledge and make decisions on risk discovery and assessment. Brainstorming meetings allow engagement teams to identify risks and initiate a free flow of ideas about how a material misstatement, whether fraudulent or erroneous, could occur (AICPA, 2012). The discussions between the engagement team members integrate various knowledges, such as documents, auditors' experience, innovative ideas, etc., which can be used for decision support to the future audit plan engagements. Accumulative knowledge in the brainstorming helps understand how more experienced auditors evaluate information and make judgements and provide compelling information as decision support for improving risk assessment. It also provides insight to regulators, practitioners and academics about auditors' risks assessment process.

To the best of our knowledge, there have been no existing researches on intelligent audit knowledge discovery and analysis from audit brainstorm meetings. The current knowledge discovery still relies on the manual process of the audio records and meeting summaries, which is burdensome, unstructured, and inappropriate for future analysis.

In this paper, we focus on extracting important brainstorming discussion contents automatically and continuously and convert extracted contents into machine-readable knowledge for future use. We propose a Natural Language Processing (NLP)-based audit plan knowledge discovery system (APKDS) for analyzing audit plan conversations and discovering knowledge. The motivation is to program machine learning algorithms to fruitfully process large amounts of natural language data from auditing documents.

There has been many issues and models studied in literature for conversation understanding, but none of them focus on audit domain scenario and what is important for audit knowledge extraction. Based on current NLP methodologies, the APKDS framework is proposed to collect and identify expertise and experience from senior auditors and specialists, and to better understand auditors' risk assessment process. Collected knowledge can be integrated to other existing CAATs as a new type of knowledge source as decision aids. The proposed tool will not replace existing CAATs but work as a supplement to them.

The structure of this paper is as follows. Section 2 introduces related work to spoken communication processing. Section 3 will introduce the proposed audit plan knowledge discovery system (APKDS) framework, and detailed design for each module will be discussed in detail in Section 4. In the end, Section 5 has a summary of this study.

3.2 Literature review

3.2.1 Recent studies in spoken content processing

Conversation is one of the most natural and efficient way of communication. There has been an important and urgent research interest in spoken content processing with increasing volumes of multimedia information in big data age (Chen et al. 2018; Bost et al. 2015). Various types of conversational setups include broadcast news, lecture recordings, voice mails and video streams, meetings, issues discussions, task assignments and planning in an organization. The contents in these conversations provide ample source material for later use (De Mori et al. 2008), such as topic identification, action item and decision detection. Audit brainstorming meetings is one example of the conversational setups which provide valuable information on how engagement teams evaluate information and make decisions.

Some important research areas in spoken conversation analysis include topic identification (Hazen 2011; Bost et al. 2015), summarization, action item and decision detection and speaker role detection. Topic identification aims at extracting main topics in a conversation. In spite of the relevant progress achieved so far, it is difficult to reliably identify multiple topics in real-life telephone conversations between casual speakers in unpredictable acoustic environments (Bost et al. 2015). Summarization aims to generate a compact and summary version of meeting discussions. These summaries can be formed by extracting original speaker utterances (extractive summarization) or by formulating new sentences for the summary (abstractive summarization) (Tur and Hakkani-Tür 2011). Prior studies showed that understanding and extracting key action items and decisions is the most common and important purpose when analyzing meetings dialogues (Banerjee et al. 2005; Lisowska 2003). Action item and decision detection aims to detect task assignments to people and associated deadlines and decision-making sub-dialogs during a meeting. These can be used to enter such information into the related person's calendar, or to track status and progress in the following meetings. Speaker role detection aims to classify each of the speakers with respect to their institutional roles, which is used in many social sciences studies. We focus on knowledge summarization in this study, which is extracting original speaker utterances and identify most important contents based on identified risk assessment topics.

3.2.2 Difficulties in spoken content processing

While speech is the most natural medium of human/human communication, little of spoken data, especially business-related data, is available for research purposes. There are two main reasons: 1. privacy and competitive advantage requirements for business meeting data and copyright issues; 2. signal quality related issues, such as non-ideal recording conditions (Tar et al. 2010). Unlike textual communication such as emails or instant messaging, almost all of spoken interactions are lost unrecorded and unprocessed (Tur and Hakkani-Tür 2011).

There are some well-known spoken meeting data used in existing studies. Switchboard (Godfrey et al. 1992), sponsored by the DARPA, was the most famous human conversations corpus. It is a large multi-speaker corpus with 2430 telephone conversations by about 500 paid volunteers on pre-determined topics, totaling 240 hours of speech and about 3 million words of text. The NLP studies then extended to two-party conversations such as multi-party conversations (or meetings), lectures, and projects that were initiated at CMU (Burger et al. 2002) and ICSI (Janin et al. 2004). More recently, several large government-funded research projects started. The AMI (Augmented Multi-Party Interaction) Consortium project and the DARPA-funded CALO (Cognitive Assistant that Learns and Organizes) project focus on conference room meetings that under controlled experimental environment (Tar et al. 2010). The CHIL (Computers in the Human Interaction Loop) project collected lectures dominated by one presenter with shorter question-answer portions and collected some “interactive” lectures involving smaller groups (Waibel et al. 2010; Tur and Hakkani-Tür 2011).

While such conversational interactions are so common, there is still no globally adopted automated or semi-automated mechanism for tracking conversations, saving the conversation content for later use, or automatically extracting certain content related features such as topics discussed or argued and decisions made (Tur & Hakkani-Tür 2011). The increased prominence that searching has become a basic user activity have shown that automate browse, summarize, and graphically visualize various aspects of the spoken content have become more and more important.

Although many studies investigated multi-party conversation analysis, no study has investigated spoken data in auditing domain meeting scenario, and no developed systems or framework in NLP literature that can process audit conversations and identify discuss contents or identify knowledge in risk assessment related conversations.

3.3 Proposed framework of continuous knowledge collection and management system

3.3.1 Objective and Motivation

Auditing is a knowledge-based service in which the management of related expertise is very important (Davenport, 1997). Auditors discussions during brainstorming meetings involve various topics on risks areas and how they achieve decisions. Current knowledge discovery in audit domain relies on manual process of audio records and meeting summaries. No developed systems or framework in NLP literature that can process audit conversations, and no existing audit tools have been developed on intelligent audit risk assessment knowledge discovery.

Despite the valuable information from the brainstorming meeting, it is very challenging to retrieve the knowledge from the audio streaming recorded during the discussion. First of all, it is difficult to extract the main contents, the key sentences that contain the most important information, from a long and unstructured audio streaming. Secondly, we need to discover the knowledge from the main contents for future analysis to support audit plan. The “knowledge” in auditing area, however, is astoundingly complex, diverse, and sometimes nonstandard, which make the knowledge discovery difficult in auditing data.

Below are the main research questions of this study:

- How to collect and extract main contents from long, unstructured spoken conversations?
- How can identified knowledge be used as decision support for future engagement cases?
- How to make this process automatically and continuously?

3.3.2 Main functions of the proposed system

Prior studies have shown that the expertise and experience from senior auditors during audit plan meetings can provide valuable knowledge for identifying and assessing risks of the entity and deciding following procedures. The proposed audit plan knowledge discovery system (APKDS) aims at achieving four main functions:

- Preprocess conversation texture data during brainstorming meetings;
- Extract and Summarize important discussion contents and convert it into standardized knowledge for information retrieval support;
- Recommend related discussion topics to auditors during brainstorming discussions;
- Allow knowledge updates when more users interact with the system.

To achieve these functions for audit knowledge discovery, NLP technologies on speech recognition, sentiment measurement and text analytics provide the practical technical support.

3.3.3 Workflow of the proposed system

Figure 10 shows the workflow of the proposed system. There are two main parts:

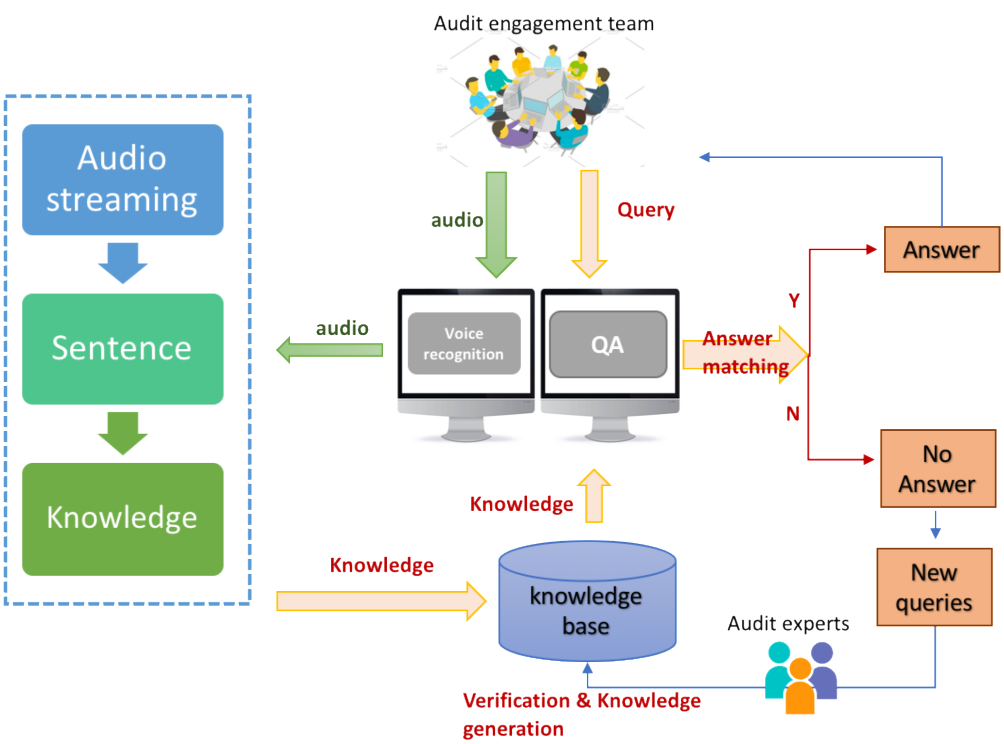


Figure 10 Workflow of The Proposed System

The first part (left side) is developed for collecting auditors' conversation and transfer it to usable knowledge. To start, the engagement team's conversation will be automatically recorded through the voice recognition processor in the system. While they are talking, their audio streaming is collected and stored in the system waiting for processing. Then NLP modules will first convert the audio into sentences, and then classify the text by identifying topics in the conversation. In the end, the conversation will be transferred into usable knowledge and stored in a knowledge base, which can be extracted and integrated with other information sources to provide decision support to auditors. The detailed NLP modules and how they work in the knowledge collection process will be introduced more in the following section.

The second part (right side) of the system is developed for auditors to ask questions. The users can interact with an intelligent question-answering interface by asking questions. Auditors ask questions to the system, then the QA module will extract the keywords in the query and look for best answer. If a matching answer is found, the answer will be given to the auditor, and if not, the query will be saved in a separate file. The topic and the value of the query will be verified by audit experts before this new topic become "officially" accepted as a topic type in text classification. Details of the answer matching process will be introduced more later.

3.3.4 Introduction of modules

Figure 11 and Figure 12 shows the audit information processing steps and related NLP modules for the Figure 10 workflow, which consists of five core modules: automatic speech recognition, streaming text segmentation, intelligent topic discovery (sentence topic discovery, topic linkage analysis), intelligent knowledge discovery, and system QA and system update.

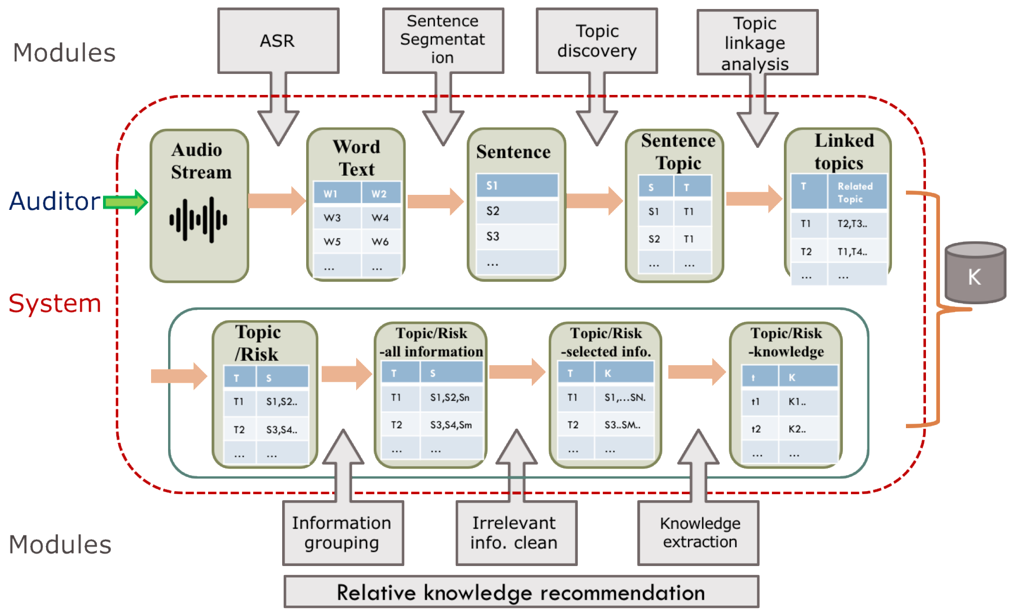


Figure 11 Modules of The Proposed System (1)

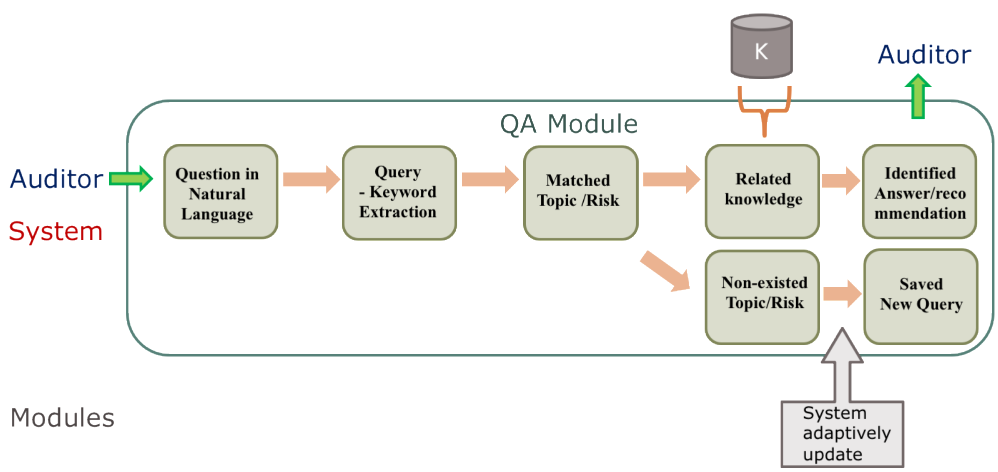


Figure 12 Modules of The Proposed System (2)

a. Automatic Speech Recognition (ASR)

The process of speech recognition is to transfer the audio speech from audit engagement teams' brainstorming meeting conversations into text. The audio is converted into steam of words through ASR. There are no punctuation marks in the initial output of

ASR. The first two boxes in Figure 11 shows the recorded brainstorming dialogue audio stream is converted into word steam if we define each word as $W, \{W_i, i = 1,2,3, \ldots n\}$.

b. Sentence segmentation

Since the output of an ASR system is a stream of recognized words without any punctuation signs (Dalva et al. 2017), the sentence segmentation (also called text streaming segmentation or dialog act segmentation) module is built to segment the text series into individual sentences ($S, \{S_i, i = 1,2,3, \ldots n\}$). Identifying appropriate sentence boundaries is an important basic for further analyzing the procedures and risks discussed during the brainstorming sessions.

c. Intelligent topic discovery

The series of segmented sentences is then analyzed to discover audit risk assessment knowledge. This module consists of two parts: sentence topic discovery and topic linkage analyzer.

- Topic discovery

Through this process, the main topic of each sentence is discovered. The brainstorming meetings cover various discuss topics on inherent and fraud risks. Each segmented sentence (S_i) will be classified to one or more topic categories (T_i), which prepares each dialogue sentence for further analysis in the next step.

- Topic linkage analysis

The topic linkage is built according to the sequence of topics covered by a series of sentences. As a result, for each important audit risk, its related risks will be discovered and the hidden relationships of different risks/topics will be identified.

d. Intelligent knowledge discovery

Based on identified sentence topic category, sentences tagged with the same topic(s) will be grouped and then irrelevant information will be removed from the grouped information. For each risk or discuss topic, the most relevant and important knowledge (words, phrases, or sentences from original grouped conversation sentences) will then be extracted, and this knowledge is the final information that will be given to future users (auditors) in their information retrieval requests.

e. System Q-A and System update (New topic collection)

As shown in Figure 12, auditors can ask questions in natural language through the QA module. The QA module will first extract the keywords of the query and match to the most relevant pre-defined topic/risk category, which helps the module to find the best answer from the knowledge base. System update module is developed to detect new topics that were not included in the existing topic list and prepare the system to collect new knowledge for the new topics (Figure 12). As a result, the system is adaptively updated for future audit audio analysis.

3.4 Modules of the proposed system

Here in this section, detailed methodology design for each part of the proposed system will be introduced: ASR and sentence segmentation, topic discovery, topic linkage analysis, intelligent knowledge discovery, system Q-A and system update. In this section, the dataset used in the module examples is sample data selected from an experimental audit brainstorming meeting audio recording introduced in Chapter 2.

3.4.1 ASR

Automatic speech recognition (ASR) is the first step in NLP. ASR is a challenging task due to the complexity of human language and the quality of conversational speech signals. There has been great progress in the automatic speech recognition (ASR) systems in recent years (Xiong et al. 2017; Shivakumar et al. 2018). Many commercial services have been developed based on current studies on speech-to-text conversion, such as IBM Watson Speech to Text service[1], Google Cloud Speech API[2] and Microsoft Custom Recognition Intelligent Service (CRIS)[3]. These services leverage machine intelligence technologies (advanced deep learning neural network algorithms) to transcribe the human voice with high accuracy. Example applications of these services include voice control of applications, embedded devices, vehicle accessories, and recording meetings and conference calls. Latest tools can convert speech of different languages into text. For example, IBM Watson Speech to Text service supports eight languages[4], and Google Cloud Speech API recognizes over 110 languages and variants.

---

[1] https://www.ibm.com/watson/developercloud/speech-to-text.html

[2] https://cloud.google.com/speech/

[3] https://cris.ai/

[4] Arabic, English, Spanish, French, Brazilian Portuguese, Japanese, and Mandarin

As a foundation of our proposed system, the performance of ASR will significantly affect the performance of other modules. The main features of current ASR technology make the objective of collecting auditors' professional knowledge possible. Firstly, it converts auditor dialogue audio into text in real-time and continuously, and retroactively updates a transcription while speaking. Secondly, it provides various interfaces that make it work for any application, which allows it to be integrated into different CAATTs when necessary.

Thirdly, it can distinguish different speakers using a single microphone. In the audit barnstorming meetings, it is also important to identify the speakers of each speech sentences. Expertise and knowledge on risk assessment from senior auditors and experts are more valuable and it is more important to collect their speech parts for further processing. The current speech to text services have some capabilities in identifying speakers. IBM Watson Speech to Text service provides the function for identifying different speakers. There have been studies on speaker role detection methodologies. To distinguish speaker role in the multi-party conversation, recent studies used turn-taking features and lexical features to cluster the roles of speakers (Yaman et al. 2010; Hutchinson et al. 2010). With speaker segmentation, the steam of words is grouped into blocks according to the role of the person in the engagement team.

Since audio conversion is not a domain specific task, in the proposed system, developers can complete the ASR task by selecting from the existing ASR services such as IBM Watson Speech Recognition instead of developing a new module. Figure 13 shows an example of the ASR converting process using IBM Watson Speech to Text tool. The input is recorded audio of the audit brainstorming conversation between the partner and manager from the experimental brainstorming case data, and the output that the tool gives automatically.

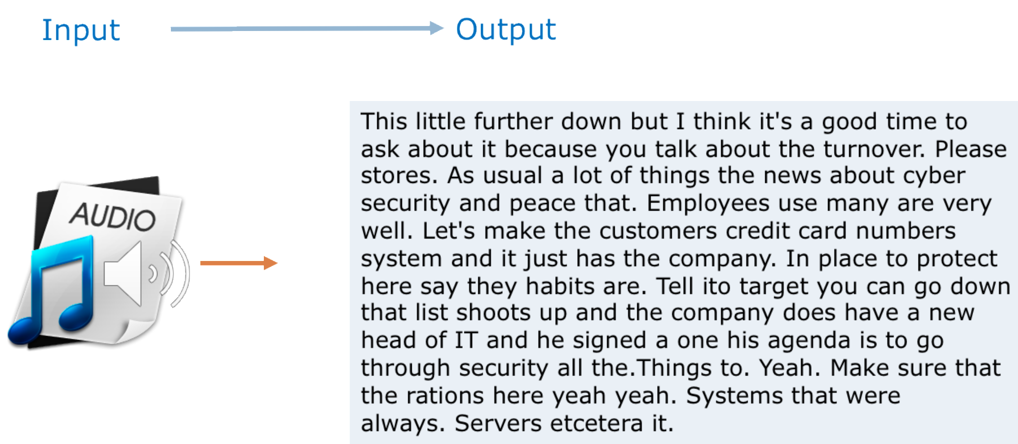


Figure 13 Output of ASR in IBM Watson Speech to Text

However, since existing ASR technology has high requirement for conversation recording environment, the converting results may not be good enough for further content analysis, e.g., the converted text with the IBM tool in Figure 13. Thus, in the proposed solution, before the next step, the texts should be adjusted manually for module training purposes. In addition, a better audio recording environment (e.g., with noise cancellation techniques) will effectively improve recording quality, e.g., preparing each audit engagement team member a nearby microphone. Figure 14 shows an example of expected converting results that will be appropriate for future content analysis after manual adjustment.

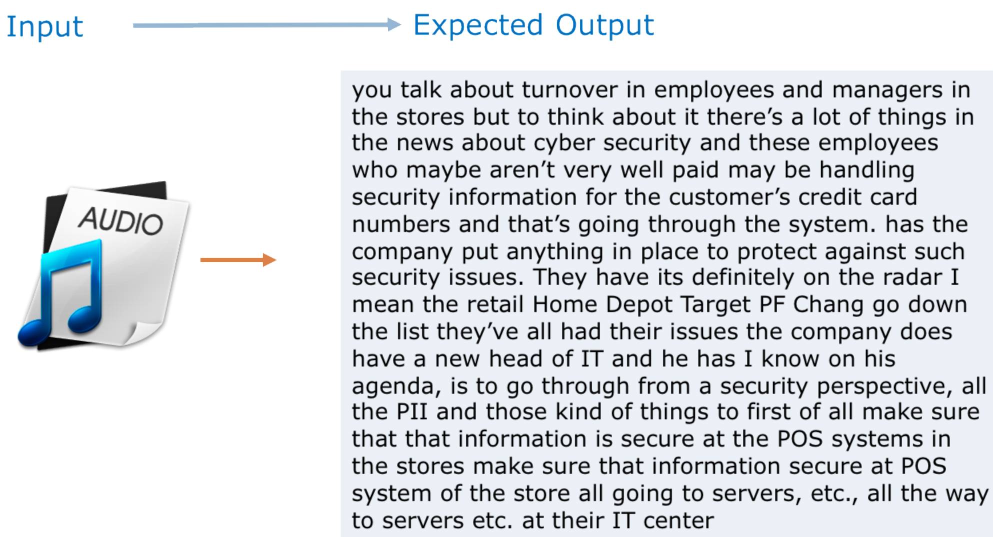


Figure 14 Expected Output of ASR

3.4.2 Sentence segmentation

Sentence segmentation process (Dialog act segmentation) is the next step and is conducted to segment the incoming audio steaming into individual sentences by determining the sentence boundaries of a stream of words (Dalva et al. 2017). Theoretically, sentence segmentation problem is actually a word boundary classification problem (Tur and Hakkani-Tür 2011), with the goal of finding the most likely boundary tag sequence. It is vital for follow-up tasks such as summarization (Liu and Xie 2008), information extraction (Favre et al. 2008), and translation (Matusov et al. 2007). The purpose of this proposed system is to collect knowledge and experience from senior auditors about clients, thus it is critical for the system to be able to accurately tag the topic of what different auditors said in each sentence. Identifying appropriate sentence boundaries is an important basic for further analyzing the procedures and risks discussed during the brainstorming.

Numerous studies showed that the lack of sentence boundaries is confusing both for humans and machines (Dalva et al. 2017). Absence of sentence boundaries can lead to meaning ambiguity for some utterances. For example, for a stream of words like “no revenue is increasing”, there can be the two possible interpretations with completely different meanings. One of them is “No revenue is increasing” and the other one is “No. revenue is increasing”.

Discriminative, generative or hybrid models have been proposed to solve this problem (Dalva et al. 2017; Tur and Hakkani-Tür 2011). One of the well-known generative approaches called hidden event language models (HELMs) were proposed in (Stolcke and Shriberg 1996).

When designing and developing the tools the analyze the conversations in the audit plan brainstorming meetings, one method that developers can use is to choose from various existing software or packages to the recorded dialogues. One type of tools uses its built-in NLP model to identify the linguistic relationships of each word and then find the best full stop of a sentence, such as the Sentence Boundary Detection (SBD) model in the spaCy (open source Python package) tool[5]. Another type of tools is the existing service such as IBM Watson Speech to Text service and Google Cloud Speech API. These services segment sentences based on the time gap of the speaker in the dialog. A punctuation mark is added in a sentence when a speaker stops talking for a short while.

Figure 15 shows an example of expected sentence segmentation based on the output of texts in Figure 14.

[5] https://spacy.io/

Input

Expected Output

you talk about turnover in employees and managers in the stores but to think about it there's a lot of things in the news about cyber security and these employees who maybe aren't very well paid may be handling security information for the customer's credit card numbers and that's going through the system. has the company put anything in place to protect against such security issues. They have its definitely on the radar I mean the retail Home Depot Target PF Chang go down the list they've all had their issues the company does have a new head of IT and he has I know on his agenda, is to go through from a security perspective, all the PII and those kind of things to first of all make sure that that information is secure at the POS systems in the stores make sure that information secure at POS system of the store all going to servers, etc., all the way to servers etc. at their IT center

1	you talk about turnover in employees and managers in the stores
2	but to think about it, there's a lot of things in the news about cyber security
3	and these employees who maybe aren't very well paid may be handling security information for the customer's credit card numbers and that's going through the system,
4	has the company put anything in place to protect against such security issues?
5	They have, its definitely on the radar, I mean the retail Home Depot, Target, PF Chang, go down the list they've all had their issues
6	the company does have a new head of IT
7	and he has, I know on his agenda, is to go through from a security perspective, all the PII and those kind of things
8	to first of all make sure that that information is secure at the POS systems in the stores
9	make sure that information secure at POS system of the store all going to servers, etc., all the way to servers etc. at their IT center.

Figure 15 Example of Expected Output of Sentence Segmentation

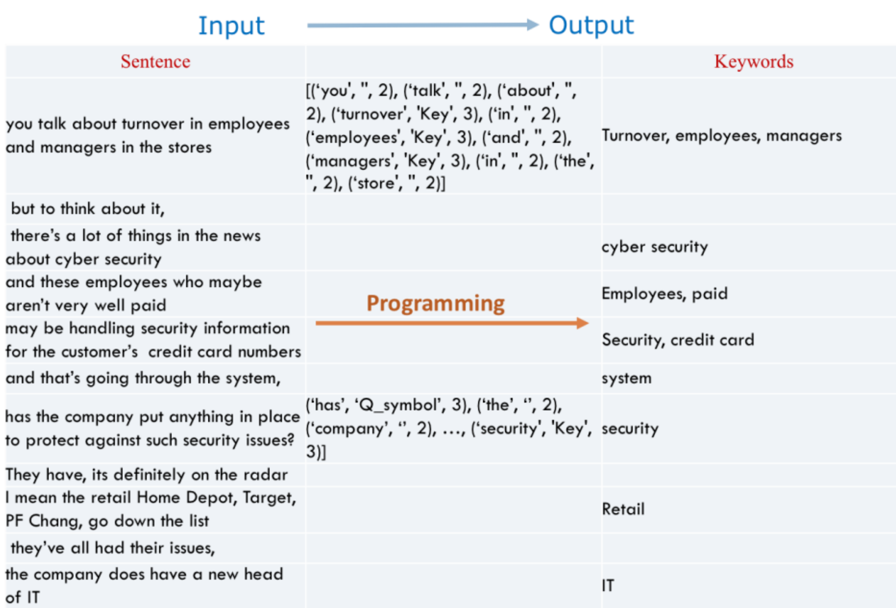
3.4.3 Intelligent topic discovery

3.4.3.1 Sentence topic discovery

In topic discovery literature, algorithms are developed to discover topics based on text documents without knowing any potential topics of the conversation. In the audit brainstorming cases, auditors discuss the risk areas and factors around the entity, the topics and expressions are more restricted than open conversations. Therefore, unsupervised topic discovery algorithms do not fit in our situations. The objective of topic discovery in this study is to identify the main topic(s) of each segmented sentence based on pre-defined risk areas/topics list. Although we pre-define some common topics for system building and training purpose, there will be extended different topics and relationships as more and more brainstorming meetings are recorded on engagement cases for different clients and industries. Thus, our proposed method can continuously enhance its expertise knowledge base.

A practical purpose of analyzing the audit brainstorming conversations is to extract key information related to risk assessment and related audit decisions. The identified risks and decisions will be used for further knowledge extraction and related topic recommendations to auditors in the future engagement cases.

In this proposed module, a three-step methodology is developed to classify each sentence into appropriate topic categories. Firstly, keywords of each sentence will be extracted through machine learning methodology. Figure 16 below shows that the example sentences in Figure 15 will be analyzed and the keywords of each sentence will be identified.



Sentence		Keywords
you talk about turnover in employees and managers in the stores	[('you', '', 2), ('talk', '', 2), ('about', '', 2), ('turnover', 'Key', 3), ('in', '', 2), ('employees', 'Key', 3), ('and', '', 2), ('managers', 'Key', 3), ('in', '', 2), ('the', '', 2), ('store', '', 2)]	Turnover, employees, managers
but to think about it,		
there's a lot of things in the news about cyber security		cyber security
and these employees who maybe aren't very well paid		Employees, paid
may be handling security information for the customer's credit card numbers		Security, credit card
and that's going through the system,		system
has the company put anything in place to protect against such security issues?	('has', 'Q_symbol', 3), ('the', '', 2), ('company', '', 2), …, ('security', 'Key', 3)]	security
They have, its definitely on the radar		
I mean the retail Home Depot, Target, PF Chang, go down the list		Retail
they've all had their issues,		
the company does have a new head of IT		IT

Figure 16 Keywords Identification

Secondly, keywords of each sentence will be matched to different "topic" based on a predefined topic-keyword dictionary. A pre-defined topic-keyword list needs to be established first which includes "topics" and "keywords" for each topic. The middle table in Figure 18 shows an example of topic-keyword dictionary. A risk topics and related keywords list is required in actual module development, which will be described more in Chapter 4.

If the keywords of one sentence can be mapped to more than one topic, then classification follows the following rules (Figure 17).

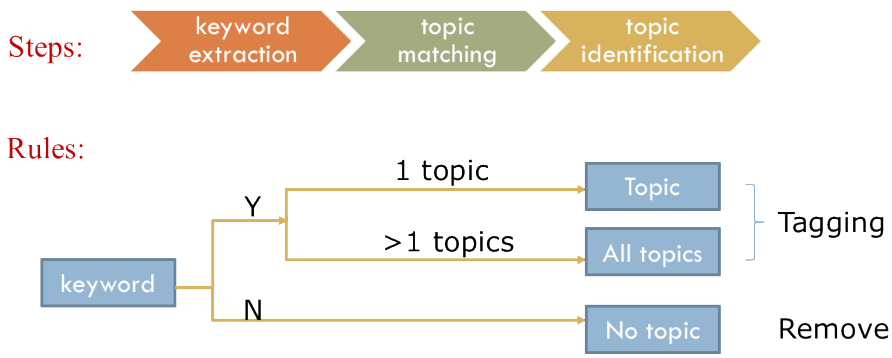


Figure 17 Sentence-Topic Classification Rules

Each sentence is then classified to one "Topic", and there are three situations:

In the first situation, there are pre-defined keywords that can be identified in the sentence. For example, in the sentence "you talk about turnover in employees and managers in the stores", keywords "turnover", "employees", and "managers" can be found, and the rule is to classify the sentence to topic "management".

In the second situation, there is no keywords can be identified in a sentence. For example, there is no pre-determined keyword in the sentence "they've all had their issues". If this happens, the sentence will be removed without classifying it to a "topic".

In the third situation, the keywords in some sentences are not from only one "topic", and there will be more than one "candidate topics". In this case, the "candidate topics" are called "topic indicators". Then we will tag the sentence with all its "topic indicators", and these results will be used for identifying the relationships of risks/topics in the following module.

Figure 18 shows the matching of the keywords from sentences to their "topic" categories. As a result, each sentence is classified to one or several topic(s).

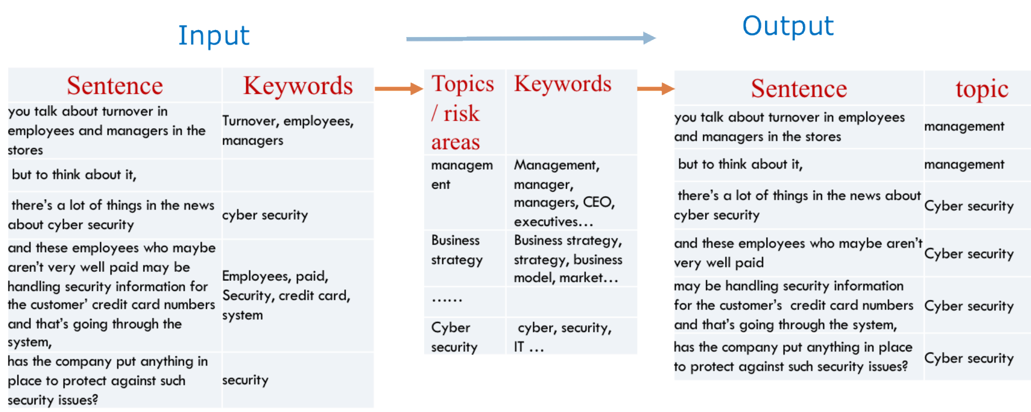


Figure 18 Topic Discovery Process

3.4.3.2 Topic linkage analysis

The topic linkage is built according to the assigned topic(s) of the sentences in the last step. In general, some risk topics are related (e.g., weather risk and geological risk) and they may be discussed in different meetings following similar logical sequence. To recommend "related topics" to the engagement team with the purpose of reminding them to discuss related risks, we build a topic transferring matrix $\mathcal{T}$ for this module.

The element $\mathcal{T}_{ij}$ represents the topic transferring frequency extracted in our training records. Once topic j is detected after topic i discussion, $\mathcal{T}_{ij} = \mathcal{T}_{ij} + 1$. In online practice, if topic i is in current discussion, we would recommend the next topic according to the highest "next-topic" probability: $\mathrm{j} = \mathbf{argmax}_{\mathrm{j}}\, \mathcal{T}_{\mathrm{ij}}$. The results of the topic linkage analysis are related topics for each important topic, which will be used as part of the recommendations to auditors later. Figure 19 shows an example of related topic "linkage point" for Topics No. 1,2,3, and 4.

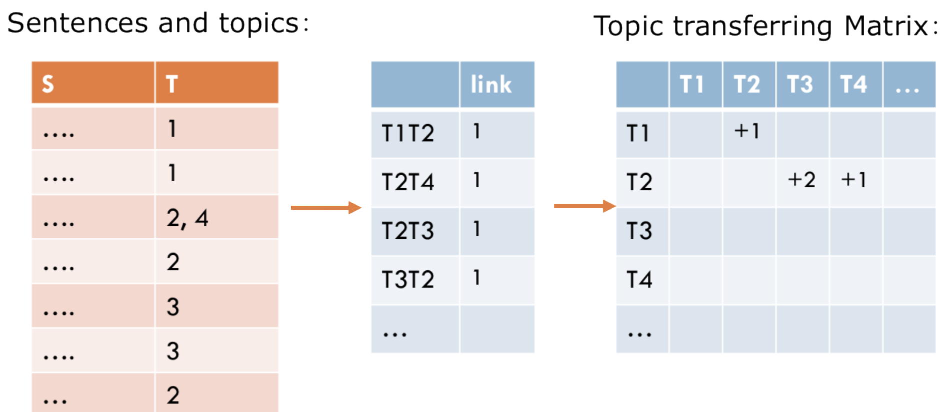


Figure 19 Topic Transferring Matrix Illustration

For example, if the module identified that when auditors discuss "IT security", they also discuss "management" before or after the topic of "IT security", then "management" will be recommended as a discuss topic to auditors when they are querying about the "IT security" during a QA interaction. Similarly, if we can identify that auditors usually discuss "geography" risks when they discuss "weather" risks, then the system will suggest auditors to think about "geography" risks. Figure 20 shows an example of established topic linkage for these two topics.

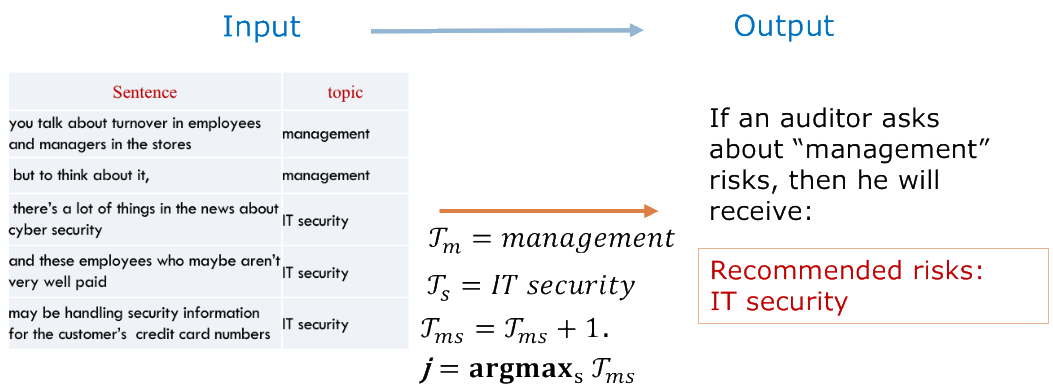


Figure 20 Example of Topic Linkage

3.4.4 Intelligent knowledge discovery

The next process is to extract most important information in numbers, words and phrases out of the grouped sentences. The reason of doing this is that the conversational information in the collected sentences could be very long and include redundant expressions, and we only want to provide auditors with the most valuable information (the knowledge) in the decision support system. The proposed module has two steps: same topic sentences grouping, and key information discovery with Dependency-Based Word Embeddings model (Levy and Goldberg 2014).

In the first step, after topic discovery process, sentences are tagged with different "topics", and they can be grouped together if they are assigned with the same "topic" tag. The grouped sentences will then be used for knowledge discovery. As mentioned in the last step, only sentences that contain keywords of that topic will be tagged and extracted, while the other sentences are considered as "not important" at this point and will be removed. However, all the sentences of each topic should be stored with a back-up, which ensures that if the extracted knowledge cannot express the full idea about the topic discussion, auditors still have the opportunity to check the original complete conversation. In addition, the back-up can be used for documentation purpose and can be integrated into other tools for potential different types of analysis in the future. This back-up can also be used for further knowledge extraction if system developers intend to improve audit knowledge extraction methodology by further defining more important knowledge from sentences (such as decision-making sentences).

In the second step, knowledge discovery aims at extracting key action items and decisions. There is some related work in detecting decision-making utterances in meetings. For example, Hsueh and Moore (2007) employed a Maximum Entropy classifier using lexical (words and phrases), prosodic (pitch and intensity), semantic (dialog act tags, temporal expressions) and contextual (relative position within the meeting) features to detect decision-making utterances. Prior studies showed that understanding and extracting key action items and decisions is the most common and important purpose when analyzing meetings dialogues (Banerjee et al. 2005; Lisowska 2003).

The representative approach for important contexts extraction is the bag-of-words approach. Compared to this method, syntactic dependencies are more inclusive and more focused (Levy and Goldberg 2014). Dependency-based method derives contexts using the syntactic relations of the words. Stanford typed dependencies (SD) representation (De and Manning 2008) is used in this method, which "was designed to provide a simple description of the grammatical relationships in a sentence" and helps people to extract textual relations easily. Some commonly used relations are "nsubj", "dobj", "prep with", "amod" etc. For example, "nsubj" means "nominal subject", which is "a noun phrase which is the syntactic subject of a clause"[6]. Figure 21 shows an example given by Levy and Goldberg (2014) on dependency context extraction.

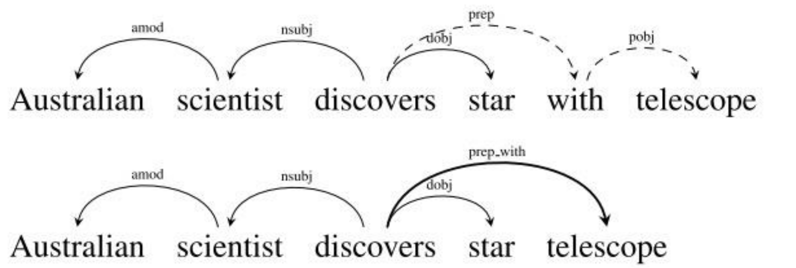


Figure 21 Dependency-Based Context Extraction Example

(Levy and Goldberg 2014)

In the dependency-based model, the SD representation is also called modifiers. For a target word W with modifiers $M_1, M_2, \ldots\ M_k$ and a head H, the model identifies the type of the dependency relation between the head and the modifier (Levy and Goldberg 2014). To extract the most important information from the collected brainstorming sentences, two things need to be defined for model training: the exact topics/risks that we want the model to work on, and what dependencies (e.g., "nsubj", "dobj", "prep with", "amod") that we think are useful to locate and extract important information out of the complete sentences.

Figure 22 shows the input and output of the knowledge discovery process. First the sentences are grouped based on their topics, and then the most important phrases or sentences will be extracted using the dependency-based method for each topic, which provides shorter and better knowledge support for auditors during brainstorming discussions. For example, the extracted knowledge for "cyber security risk" is shorter and easier to read than the original grouped sentences.

6 http://universaldependencies.org/docs/en/dep/nsubj.html

Input

Sentence	Topic
You talk about turnover in employees and managers in the stores	management
But to think about it,	management
There's a lot of things in the news about cyber security	cyber security
And these employees who maybe aren't very well paid, may be handling security information for the customer's credit card numbers and that's going through the system,	cyber security
Has the company put anything in place to protect against such security issues	cyber security

Output

Topic	Collected knowledge
management	you talk about turnover in employees and managers in the stores.
cyber security	there's a lot of things in the news about cyber security. and these employees who maybe aren't very well paid may be handling security information for the customer's credit card numbers. has the company put anything in place to protect against such security issues

Topic	Knowledge
Management risk	Turnover in employees and managers
Cyber security risk	Employees who maybe aren't very well paid may be handling security information for the customer's credit card numbers

Figure 22 Output of Intelligent Knowledge Discovery

3.4.5 System Q-A and system update

3.4.5.1 Question Answering

Based on the developed topic linkages, the system then prepares most appropriate answers to audit engagement teams given a question asked by an auditor and recommend related discuss topics (risk areas) to him/her.

As shown in Figure 23, when a question is generated, the system will identify the keywords in the Question and generate a Query which contains the keywords, then the pre-defined topic-keyword dictionary is applied to help identify the topic/objective of this Question. Based on the question topic, this module will extract stored related knowledge from the knowledge base prepared in prior steps, and then return the knowledge as the answer, together with recommended "other important topics and procedures" for the engagement team to consider.

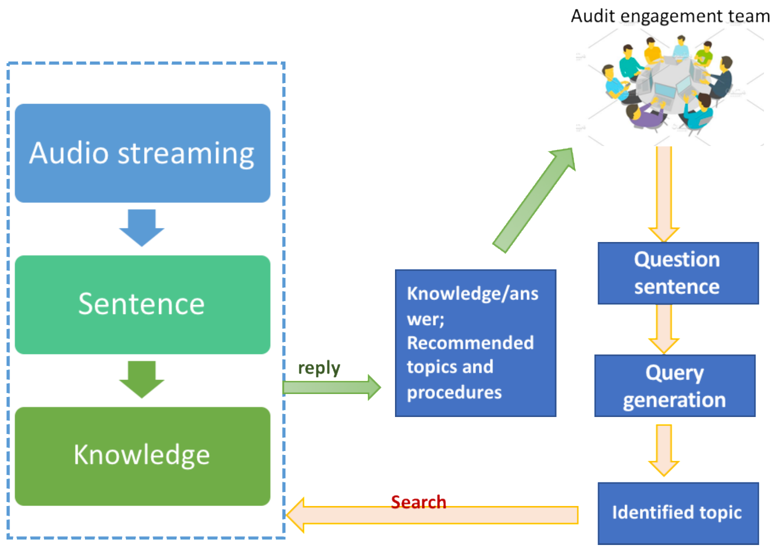


Figure 23 Question Answering Module

An answer contains three parts:

1. Extracted knowledge from the Intelligent Knowledge Discovery module;

2. Related topics identified in the Topic Linkage Analysis module;

3. Recommended risk areas and procedures based on pre-defined recommendation table.

In the second part, related topics identified in the topic linkage analysis module will be recommended as “related topics” to the auditors.

In the third part, if an audit firm has their own pre-defined templates/ checklists on brainstorming meetings, the related risk areas and procedures that are required by the audit firm as discuss topics will be recommended to auditors as well when they are querying a “topic” in the QA query. A recommendation table needs to be created for the system module to find the recommendations.

Figure 24 shows an example question that could be asked by an auditor and the possible answer that he/she could receive.

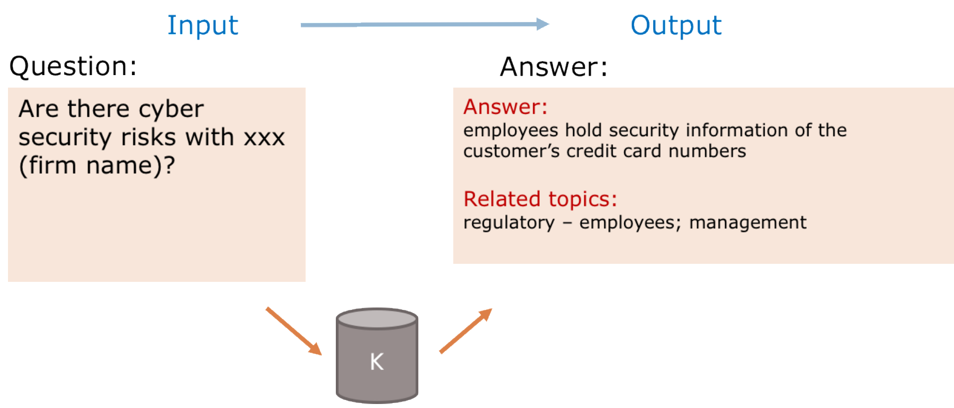


Figure 24 Example Answers to A Question

3.4.5.2 System adaptively updates

As shown in Figure 25 (the box at the lower right corner), the system updates module aims at identifying and saving new knowledge. New topics that were not covered in existing topic list can appear in a user QA interaction. When the QA module cannot identify what topic/risk area a question is asked for, the system update module will consider this query a potential new topic/risk area and store this query separately in a file. Audit experts will be needed to manually review and verify these new topics before they can be updated permanently in the knowledge base topic list. Then the system will be able to re-analyze existing brainstorming discussion documents to discover knowledge and also to collect and analyze future conversations to improve the knowledge base for this topic/risk. As a result, the QA module is adaptively updated for future usage. Experts should also review the newly collected knowledge in this module before it is updated into the knowledge base. Figure 26 shows an example of the situation when a new topic is asked and there has been no related knowledge prepared in the knowledge base.

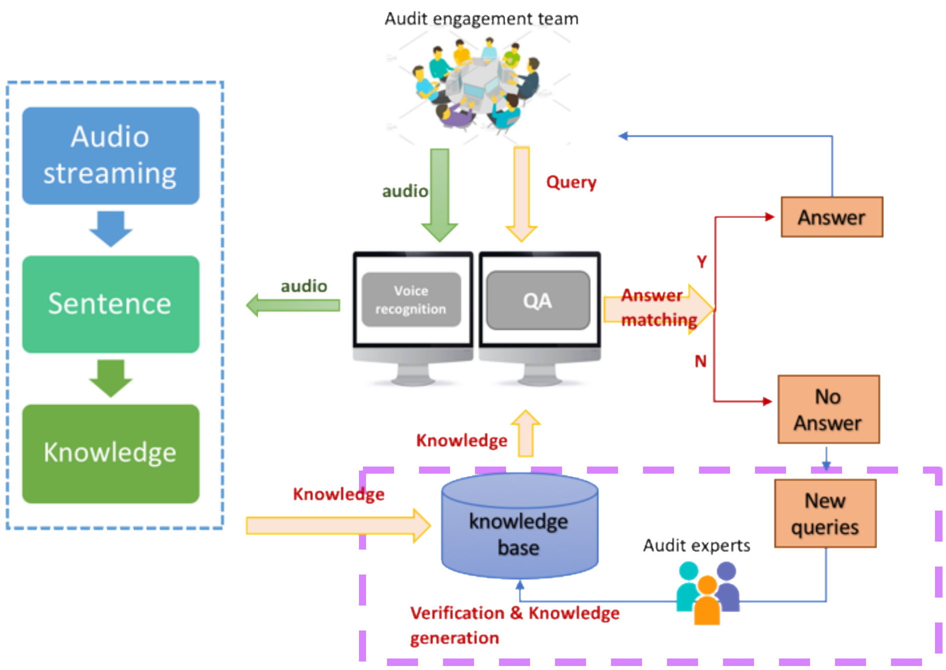


Figure 25 System Adaptively Updates Module

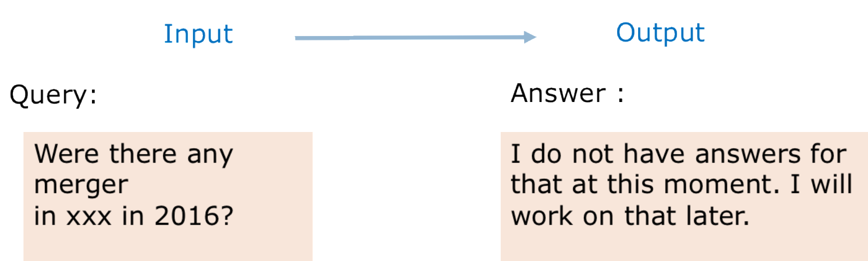


Figure 26 Example of New Topic in Auditor Query

3.5 Summary

In conclusion, this paper proposes an audit plan knowledge discovery system (APKDS) to collect and analyze senior auditors' professional knowledge and expertise during audit plan brainstorming meeting dialogues. As a result, the audio conversations can be continuously transferred into standardized information in the knowledge base which can support future engagement teams during brainstorming discussions.

From system evaluation perspective, according to Peffers et al. (2012) classification of artifact types in design science research, we can classify the proposed NLP audit conversation analysis framework as an instantiation. Instantiation is defined as “the structure and organization of a system’s hardware or system software or part thereof”. The reason of the classification is that the designed idea can intendedly be expressed by an example system, rather than in a modeling language. Prior literature showed that prototype is a common method for instantiation evaluation (Chen et al. 2002). Using a prototype can “provide strong evidence when used to show that a design works as intended, is useful for its intended purpose, or has the potential to achieve an expected performance level” (Peffers et al. 2012). Therefore, a porotype will be proposed in Chapter 4 to help evaluate this framework.

CHAPTER 4: PROTOTYPE OF AN NLP-BASED AUDIT CONVERSATION ANALYSIS AND KNOWLEDGE EXTRACTION SYSTEM

4.1 Introduction

Chapter 3 proposed an audit plan knowledge discovery system (APKDS) for continuously collecting and analyzing audit plan brainstorming discussions. The main knowledge extracted from the discussions are part of the knowledge database in the proposed cognitive assistant to support future engagement cases. This framework contributes to both academic research and practice. The extracted knowledge on how auditors identify risks and make audit decisions provide insights to researchers who study auditor behavior and decision-making related topics. However, despite the significant benefits from the audit plan discussion analysis, the development of such an intelligent audit conversation analysis system remains challenging. First of all, the fundamental of our proposed system is a Natural Language Processor which must be able to correctly identify key words from auditing conversation scenarios. Further analyses of the main contents, such as the topic-topic analysis and important knowledge extraction, are machine learning tasks.

Indeed, the emergence of Natural Language Processing (NLP) and machine learning techniques enable a new paradigm for enhancing the speech recognition and speech topic discovery. However, due to the insufficient knowledge database, there is not enough attention paid to audit plan discussion analysis. In this paper, we integrate NLP and machine learning techniques to propose a prototype for the APKDS framework and illustrate the development of important models using Python[1] and the NLP platform spaCy[2] (open source Python package). The purpose of this study is to propose how existing NLP platforms can potentially be utilized by audit firm and system developers to build the important modules for an audit conversation knowledge discovery system. Although we do not build a complete working prototype in this paper, this is the first attempt of illustrating the development of an NLP-based intelligent system for auditing plan knowledge collection and analysis. Here, we select sample data from the experimental audit brainstorming cases to demonstrate the functionality of the proposed prototype models. The proposed prototype achieves three main objectives. Firstly, it provides important topic discovery and related topic recommendation to auditors for auditors' decision support. Secondly, it can extract main contents from brainstorming discussions and transfer them into knowledge for future audit plan engagements. Thirdly, its semi-supervised adaptive learning capability ensures that the prototype can automatically enrich its knowledge database.

The structure of this paper is as follows. Section 2 provides a literature review on current popular NLP platforms/software. Section 3 describes the construction of audit knowledge database and the source of dataset used in the development of our models. Section 4 introduces the training and testing datasets. The fundamental open-source software of the prototype - spaCy is introduced in Section 5. Then, Section 6 proposes the detailed design of all the models in the prototype and shows experimental results of the

[1] https://www.python.org/

[2] https://spacy.io/

training and testing process. Section 7 . Section 8 introduces the proposed application of the cognitive assistant to other audit phases, and Section 9 is the summary.

4.2 Literature review

Since this study aims at proposing how existing NLP platforms can potentially be used to build the NLP-based audit plan knowledge discovery system, we first look into the NLP platforms that are available. Recent years have witnessed an increasing number of accurate and fast NLP platforms based on dependency parsing, and most of them are publicly available and easy to use. Dependency parsing is an approach to automatic syntactic analysis of natural language based on the theoretical linguistic tradition of dependency grammar (Kübler et al. 2009). It is now widely applied in NLP studies to build language analysis models.

A dependency parser will analyze the grammatical structure of a sentence and establish relationships between "head" words and “dependent” words. A dependency relation holds between a syntactically subordinate word, called the “dependent”, and another word on which it depends, called the “head”. Figure 27 below illustrates a dependency structure of an English sentence, and the dependency relations are represented by arrows pointing from the head to the dependent. Each arrow in Figure 27 has a label which indicates its dependency type. For example, the noun “news” is a dependent of the verb “had” with the dependency type Subject (SBJ). The noun “effect” is a dependent of type Object (OBJ) with the same head verb “had”. Also, the noun “news” is a head in the relation to the word “Economic”, which has the Attribute (ATT) relation to its head (Kübler et al. 2009).

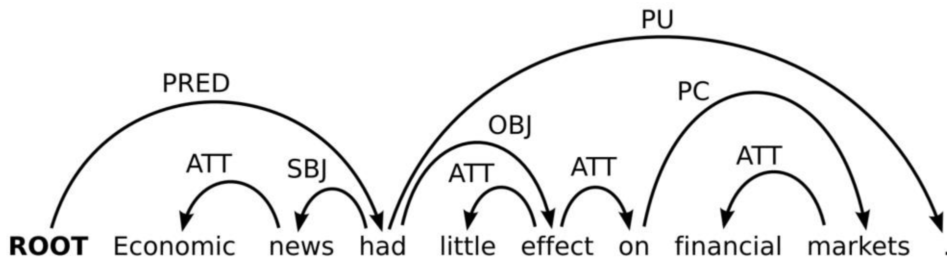


Figure 27 Dependency Structure for An English Sentence

(Kübler et al. 2009)

Since there have been different dependency parse-based NLP platforms available, we need to compare these platforms and select one that is most appropriate for the prototype development. Choi et al. (2015) summarized the state-of-the-art NLP applications using dependency parsing, as shown in Table 8.

Parser	Approach	Language	License
NLP4J[3]	Transition-based, selectional branching (Choi and McCallum 2013)	Java	Apache
Mate-Tool[4]	Maximum spanning tree, 3rd-order features (Bohnet 2010)	Java	GPL v2
RBG[5]	Tensor decomposition, randomized hill-climb (Lei et al. 2014)	Java	MIT
NNDEP[6]	Transition-based, word embeddings (Chen and Manning 2014)	Java	GPL v2
Turbo[7]	Dual decomposition, 3rd-order features (Martins et al. 2013)	C++	GPL v2

[3] https://emorynlp.github.io/nlp4j/

[4] https://code.google.com/archive/p/mate-tools/wikis/ParserAndModels.wiki

[5] https://github.com/taolei87/RBGParser

[6] https://nlp.stanford.edu/software/nndep.shtml

[7] http://www.cs.cmu.edu/~ark/TurboParser/

spaCy[8]	Transition-based, greedy, dynamic oracle, Brown clusters	Cython	Dual

Table 8 Current Open Source NLP Applications

(Choi, Tetreault, & Stent 2015)

The parser NLP4J (Choi and McCallum 2013) is a toolset written in Java for Natural Language Processing. It is built based on selectional branching which uses confidence estimates to decide when to employ a beam search. Mate-tool parser (Bohnet 2010) provides the following technology: transition-based dependency parser, beam-search and early update, graph-based completion model, joint Part-of-Speech tagging, joint Morphologic tagging, and Hash-Kernel. It uses the passive-aggressive perceptron algorithm as a Hash Kernel, which substantially improves the parsing times. RGB (Lei et al., 2014) was developed by the NLP group from Massachusetts Institute of Technology. It contains a Java implementation of a syntactic dependency parser with tensor decomposition and greedy decoding. The developers use tensors to map high-dimensional feature vectors into low dimensional representations. The parameters were explicitly maintained as a low-rank tensor to obtain low dimensional representations of words in their syntactic roles, and to leverage modularity in the tensor for easy training with online algorithms. NNDP (Chen and Manning 2014) uses a neural network classifier to build a greedy, transition-based dependency parser. This classifier learns and uses a small number of dense features. It can work very fast, while achieving an about 2% improvement in unlabeled and labeled attachment scores on both English and Chinese datasets. Turbo uses

[8] https://spacy.io/

AD[9] (Alternating Directions), an accelerated dual decomposition algorithm which was extended to handle specialized head automata and sequential head bigram models. AD is suitable for dealing with declarative constraints, which are convenient in NLP for expressing rich prior knowledge (Martins et al. 2013).

Among these open source platforms, prior literature has confirmed and concluded that spaCy is currently the fastest platform in the world and it achieves the most balanced performance in terms of accuracy and efficiency (Honnibal and Montani 2017; Choi et al. 2015). In this study, spaCy will be used to develop the NLP related modules for the proposed audit plan knowledge discovery prototype.

4.3 An overview of the proposed prototype framework

The prototype framework is proposed based on the audit plan knowledge discovery system (APKDS) developed in Chapter 3. Figure 28 shows the important modules proposed in the APKDS framework and how these modules can be realized by the proposed prototype models. The important APKDS modules[10] include Automatic Speech Recognition (ASR), sentence segmentation, topic discovery (keyword extraction, topic matching), topic linkage analysis, knowledge discovery, Question-Answering, and system topic update. In this proposed prototype, we use spaCy’s dependency-based algorithm to process the natural language data extracted from the experimental audit brainstorming

[9] http://www.cs.cmu.edu/~ark/AD3/

[10] Detailed descriptions of the modules are in Chapter 3

cases, and use Python programming to build models and demonstrate each proposed module.

As shown in Figure 28, in the proposed prototype: the sentence segmentation module is developed with spaCy's Sentence Boundary Detection (SBD) model; keyword extraction is realized with spaCy's dependency parser and a pre-defined sentence-keyword table; topic matching is completed with a pre-defined keyword-topic table; topic linkage analysis is built by a topic transferring probability matrix; and audit knowledge discovery, QA and system topic update are realized with spaCy's dependency parser. The procedure of converting an audit brainstorming meeting conversation from audio to text is proposed to be completed by an existing ASR tool, which will not be discussed in this paper for prototype development. Audit risk recommendations generated from the topic linkage analysis module and audit knowledge collected from the knowledge discovery module will be stored into the knowledge base. When an auditor interacts with proposed system prototype and asks questions through the QA module, knowledge base will be used to generate answers. New discuss topics/risks identified by the system topic update module will be updated to the (K, T) table after being verified by audit experts.

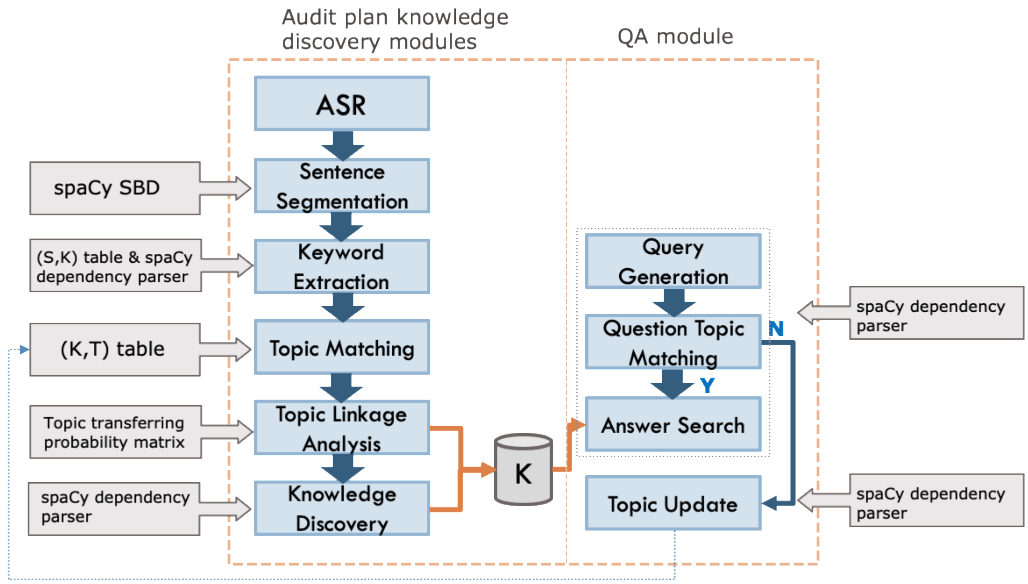


Figure 28 Overview of The Proposed Prototype Framework

4.4 (S, K) and (K, T) table construction

In order to design the NLP-and-ML based audit knowledge discovery system modules, it is important to construct a knowledge database for the system to get trained from the historical auditing discussions. This knowledge database is composed of two main tables: (S, K) and (K, T) (S represents sentence, K represents sentence keywords, and T is keyword related topics).

The (S, K) (or sentence-keyword) table and the (K, T) (or keyword-topic) table will be constructed from three sources. Firstly, in the beginning, the sentences, keywords and related topics in the tables should be created by labeling the historical audit brainstorming conversation texts. Secondly, when the system is trained and starts to work, it will collect and analyze more conversations from new brainstorming meetings, which will enrich existing knowledge base. Then, when users interact with the system through the Question-Answering module, the system will further identify new topics from the queries and update them to current topic list for future knowledge classification. With identified topics from discussions, the database will contain a variety of topics/risk areas about audit plan risk assessment.

Although the two tables are supposed to be updated and expanded continuously during prototype training and testing, the initial version of the two tables should be created for model development in this paper. The four experimental audit brainstorming meeting datasets[11] introduced in Chapter 2 are used to construct the initial (S, K) and (K, T) tables and train prototype modules. The discussion materials, main contents, referring keywords, and the topics are collected to help the system build its algorithms from the raw discussion materials. The four different brainstorming meetings in the experiment were recorded as audio files originally, and then converted into texts using existing Automatic Speech Recognition (ASR) tools. The converted experimental dialogues capture how evidence is evaluated and how audit risk assessment judgements are made in the brainstorming sessions in audit plan.

Knowledge extracted/learned from these cases provide important information for the prototype development in three main aspects:

- Framework and module improvement: the conversations in the cases help better understand the actual procedures and format of brainstorming sessions

[11] This database is used in the working paper: The Use of Verbal Protocol Analysis to Describe the Planning Risk Assessment Discussion of Audit Partners and Managers, by Helen L. Brown-Liburd, Theodore Mock, Andrea M. Rozario and Miklos A. Vasarhelyi

- Main topics/risk areas discovery: the conversations include detailed topics that were discussed during brainstorming meetings for risk identification and assessment. The topics and their referring keywords summarized from converted experimental brainstorming cases can be used to develop the initial topic-keyword table, which is a key component for the sentence topic discovery module.
- Initial QA knowledge base development: auditors' expertise and experience extracted from these brainstorming cases will be used to build the initial knowledge base for the QA module.

a. Sentence-keyword (S, K) table

To process conversational texts and extract most important information from them, the proposed system firstly needs to be trained to be able to identify the keywords of a new sentence under the accounting and auditing scenario. A (S, K) table is a system training dataset which can be created by manually tagged keywords of each sentence in the sample training data. In this study, the keywords of a sentence in the training dataset were determined based on our judgments on the importance of a word to the sentence in the audit plan scenario. Table 9 shows some example sentences selected from the experimental brainstorming cases and corresponding tagged keywords.

Sentence	**Keyword**
I have noticed they had an impairment this year for a facility they have that is out for sale	impairment
and it looks like they did a write-down of some assets over there as well	write-down, assets
so now that one is done and there are some different intangibles and what not and	intangibles
that actually speaks to the complexity of all the intangible	intangible

and long-term asset impairment analysis	long-term asset, impairment
because of the level its aggregation that you have to do in that	aggregation
they are not close to losing the investment in the plant there and	investment, plant
they do not really have many intangibles there, but it is something you got to think about	intangibles

Table 9 Sentence-Keyword Table Example

b. Keyword-topic (K, T) table

A keyword-topic list needs to be created to identify the most important topic of a sentence given the identified keywords in that sentence. Since the proposed system is developed for audit plan brainstorming meetings, the conversation topics are domain specific, which focus on inherent risk and fraud risk identification and assessment. Therefore, all important discuss topics should be pre-defined first and then knowledge can be collected and organized based on those pre-defined topic categories. In addition, the topics and keywords can be tailored and expanded based on specific needs and requirements of each audit firm.

Based on the requirement of audit standards, we summarized the most important audit plan brainstorming meeting discuss topics in Table 2-4 of Chapter 2 and classified them with three levels. For example, Table 10 below shows the topic category "Entity understanding" and its sub-categories.

Level-1 categories	**level-2 categories**
Entity understanding	Background, Business operations, Management, Investments, Financing activities, Financial reporting, Multi-locations, External factors, Entity performance measures

Table 10 Topic Categories Example

In this paper, the keywords for each topic category are composed from three main sources:

Firstly, we manually tag each sentence in the four brainstorming meeting dialogues with a topic/risk area based on our judgments, and then add the identified keywords of a sentence to the topic.

Secondly, since the four brainstorming cases only provide limited number of total sentences, we manually added some keywords for each topic that didn't appear in the four cases. For example, in the "Geography" topic, although we only found the words "northern" and "western" in the cases, we added other similar words to the topic such as "eastern" and "southern".

Thirdly, for each English word, there could be different variations such as singular and plural, passive, active, noun, adjective, adverb, different tenses such as past tense and current tense, and synonyms. We did not add these variations here in the keyword list because the NLP model we use to train the module is capable of dealing with word variations.

Table 11 shows examples of detailed design of the keyword-topic list for the topic Impairment, Weather, Management, Cyber security, Geography, Seasonality, and Regulatory – employees.

Topic	Keywords
Fixed assets/intangibles-impairment	Impairment, intangible, asset, CAPEX, Facility, write-down, long-term asset, aggregation, invest, plant, cash, valuation, acquisition, operate, investments, projections, written off

Weather	weather, hurricanes, Sandy, earthquake, flood, fire, snow, loss, insurance
Management	Operation, employee, manager, turnover, demographic, staff, executive, payroll, support Center, CEO, CFO, CTO, head, management, EPS, counsel, compensation, incentive, board, bonus, ethics, compliance, centralized, structure, relationship, culture, leadership, incented, top line, budget, policies, meetings, involvement, reputation, transparent, central group, individual, president, conservative
Cyber security	Cyber, employee, security, credit card, PII, IT, secure, POS, server, fence, privacy, control, law suit, data protection
Geography	Geographic, east, west, north, south, Southeast, southwest, south east, East Coast, West Coast, Northern, western, eastern, southern, loss, state, insurance; (all states name such as New Jersey…)
Seasonality	Seasonality, second quarter, first quarter, third quarter, fourth quarter, Spring, summer, fall, winter
Regulatory - employees	Employee, seasonal employees, part-time employees, hourly employees, wages, minimum wage, cost cutting, employee abatement, payroll, law changes, regulatory, regulation

Table 11 Example of Keyword-Topic Table

In addition to the keyword-topic list, words and phrases on question type (whether a sentence is question or a statement), industry, company name, and year (time) were also labeled as keywords in the training dataset. We call these classes of words "entity" and summarize them in the "keyword label table" (Table 12). Keyword label table is a higher-level table which is processed before the keyword-topic table. A keyword defined in the keyword-topic list above will be tagged with a "keyword-topic" label during training process first before they are used to identify the specific topic of the sentence. The following table shows some examples of the keywords under these labels.

Entity	Keywords

Question	When is, who is, where is, why does, why not, what about, what is, how about, how long, does it, do they, is that
Industry	Agriculture, Manufacturing, Wholesale, Retail …(US Census Bureau 2017)
Organization	Examples (Remove INC, CORP, LTD, -CL A, CO, -CL I, FD, FUND at name end): AAR, American Airlines Group, CECO Environmental… (WRDS Database 2018)
Year	2017, 2016, 2015, 2014
keyword-topic	(Words in keyword-topic table)

Table 12 Example of Keyword Label Table

4.5 Dataset for Training and Testing

In this section, the selected contents from experimental auditing brainstorm cases for system algorithm training and testing will be introduced.

The original experimental data are four meeting audio recordings. In order to receive better analytical result in topic identification model training and testing, the audio recordings have been converted into textual documents using existing ASR technology, and the words have been adjusted for content analysis purpose. Thus, the training and testing datasets in this study start from converted textual documents of the original audio recordings.

In order to build the prototype models, the first three of the four experimental cases are selected as training data sets, and the Restaurant case is selected as the testing data set. In the final version of a complete prototype, the whole conversations should be analyzed, and all the risk assessment discussion topics should be extracted. To start, conversations on four risk areas are selected first to build the training model and testing. The four topics are:

- Fixed assets/intangibles- impairment

- Weather
- Management
- Cyber security

All the sentences that relate to these four topics are selected from the four cases to be the training and testing datasets. In the training data, all the selected sentences are manually processed. A complete sentence-keyword table for all sentences (examples are in Table 9) and a keyword-topic table is developed (Table 11).

The total 49 sentences on the same four topics in the Restaurant case are extracted as testing data. These sentences are used to test the established models for the proposed prototype.

Table 13 shows the discussed topics in each of the case. If the meeting includes a discussion on the topic, then the box is checked as "√", otherwise it is "X". In summary, 148 sentences in the 3 cases on these four topics are used for training the model for knowledge collection, classification, and summarization.

Data type	Case No.	Topic			
		Fixed assets /intangibles-impairment	Weather	Management	Cyber security
Training	1 (Electronics)	√	√	√	√
	2 (Retail)	√	X	X	√
	3 (Rental)	√	X	√	X
Testing	4 (Restaurant)	√	√	√	√

Table 13 Extracted Four Topics in The Four Cases

.

4.6 Fundamental techniques: spaCy

The Natural Language Processing task of the proposed prototype modules is completed by spaCy, an open source Python package. This NLP platform is designed for large-scale information extraction from text sequences. It is written from the ground up in Cython[12]. The Cython language is a superset of the Python programming language that additionally supports calling C functions and declaring C types on variables and class attributes. It is designed to give C-like performance with code that is written mostly in Python. Compared to other major language processors, such as SyntaxNet, NLTK and CoreNLP, spaCy provides the best functionalities in an efficient way (Honnibal and Montani 2017). The early evaluation demonstrates the effectiveness of spaCy in text classification, with an accuracy as high as 92.6 (Choi et al. 2015).

The Figure 29 below presents the architecture of spaCy. The central components in spaCy are the Doc and the Vocab. Doc is a container that stores token data (i.e. a word, punctuation symbol, whitespace) from textual information and all their annotations. Vocab contains a set of look-up tables that make same information to be stored only once but be available across documents, which saves memory of the processor and ensures there's a single source of truth. Tokenizer is used to segment texts and create Doc objects with the discovered segment boundaries. Language is a text-processing pipeline that coordinates various components/models such as Dependency Parser and Entity Recognizer. SpaCy’s NLP models are used in sentence segmentation, keyword extraction, knowledge discovery, QA and system topic update modules for the proposed prototype.

---

[12] https://cython.org/

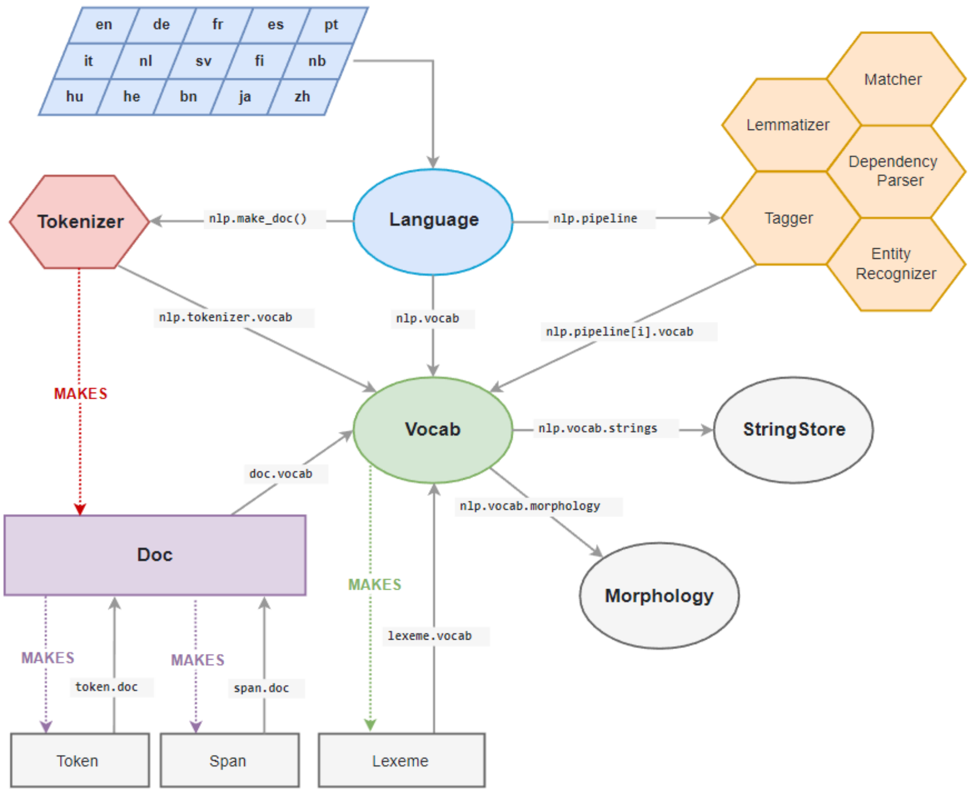


Figure 29 Architecture of SpaCy

(spaCy: https://spacy.io/api/)

4.7 Prototype development

In this section, we illustrate how to integrate NLP and machine learning techniques to develop the modules for the proposed prototype.

4.7.1 Sentence Boundary Detection (SBD)

Before audit brainstorming content analysis, firstly, an automatic sentence segmentation model should be applied to segment sentences with appropriate punctuations. Although some current Automatic Speech Recognition tools such as IBM Watson Speech to Text service can help create sentence boundaries, but the results are not good enough for further content analysis.

In this study, the Sentence Boundary Detection module from spaCy is applied the training and testing dataset to determine the boundaries of sentences. It provides an effective and efficient methodology that system developers can use to automatically segment long meeting conversations for future experimental brainstorming cases. The spaCy package extracts linguistic features of each word (called a “token” in the model) in the document, and use its built-in tokenizer to tag each token with a part-of-speech tag (e.g. VERB) and some amount of morphological information, e.g. that the verb is past tense. Then the model can identify the linguistic relationships of the tokens and in the end find the best point that separate two sentences. The SBD is implemented by analyzing both the text semantic roles and the existing punctuation between texts.

For example, given an input speech record:

text=u"the company has a new head of IT, he has a security perspective."

The SBD will first discover the semantic roles of each token and then separate them into individual sentence. The module block in our system is implemented as Figure 30:

```
#this block presents the implementation of sentence segmentation
import spacy
import random
from spacy import displacy
from spacy.tokens import Doc
nlp=spacy.load('en')
text=u"Facebook has a new head of IT. and he has a security perspective."
doc=nlp(text)
i=0
options = {'compact': True}
for span in doc.sents:
    print("sentence %d:" %i, span)
    i+=1
displacy.render(doc,style='dep',jupyter=True,options={'distance':75,'compa
ct':True})
```

Figure 30 Sentence Boundary Detection (SBD) Module

As a result, this module will output the segmented sentences.

sentence 0: the company has a new head of IT.

sentence 1: he has a security perspective.

Following figure (Figure 31) shows how the SBD module identifies dependencies of the tokens in the two sentences.

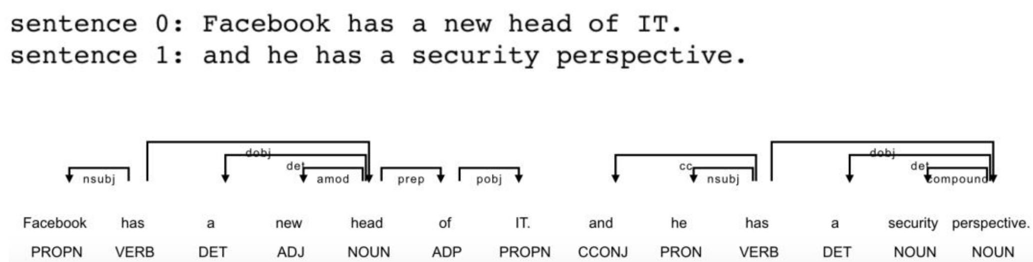


Figure 31 SBD Module Output

4.7.2 Sentence Topic Discovery

After the textual dataset is processed by the SBD model, sentences generated will be fed into the dependency parser for keywords extraction. Topic discovery is achieved by a proposed semi-supervised learning process. The semi-supervised learning for topic discovery starts from a set of labeled sentences, a set of topics, and the topic related keywords.

There are two stages in this sentence topic discovery process:

- Stage 1: Segmented training sentences are fed into the parser to extract keywords and entities
- Stage 2: Topic discovery through (K, T) table mapping

4.7.2.1 Stage 1: keyword extraction

The objective of the keyword extraction model is to train the system to correctly identify “important words” for a new sentence. For example, in the audit brainstorming, when the engagement team discusses “Apple has several new products this year”, the word “Apple” is supposed be recognized and tagged by the system as company name in this scenario. In another case, auditors may discuss a client of retail industry, and the word “apple” means the fruit instead of a company.

In this prototype, we treat the keyword discovery as a text classification problem which is achieved by spaCy’s keyword discovery model: Bidirectional Long Short-Term Memory (B-LSTM) networks. This algorithm is used for processing sequential data. This supervised learning method trains a special recurrent neural network to capture long-term sentence dependencies. It can achieve a character-level 92% prediction accuracy (Zhang et al. 2015).

The training process is illustrated in Figure 32. The “Training data” represents the existing sentence-keyword (SK) table (such as examples in Table 9). The “text” and its “label” are the current labeled sentence. By parsing the labeled sentence, a “Doc” is created to update the model. The “optimizer” will handle the status of model updating. The labeled sentences will mark all important tokens, including entities (such as Companies, Organizations, etc.), locations (including Cities, Provinces, etc.), time (including Month, Year, etc.), and topic related keyword (including tax, labor cost, impairment, etc.).

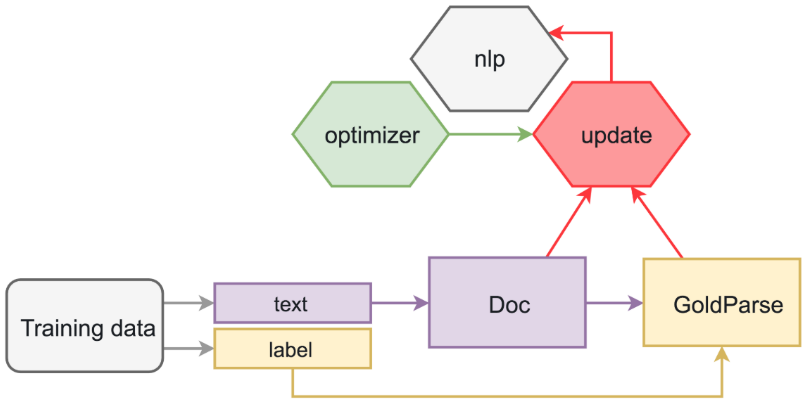


Figure 32 Spacy Keyword Discovery Training Process

(spaCy: https://spacy.io/api/)

The Figure 33 below shows some example tags of identified keywords and phrases according to the pre-defined SK table. For example, “Home Depot” is the name of a company (organization), thus it is extracted and tagged as “ORG”; “cyber security” and “head of IT” are keywords/phrases that were defined in the topic-keyword list, thus they are tagged as “keyword” which will later be used to match to the related topic.

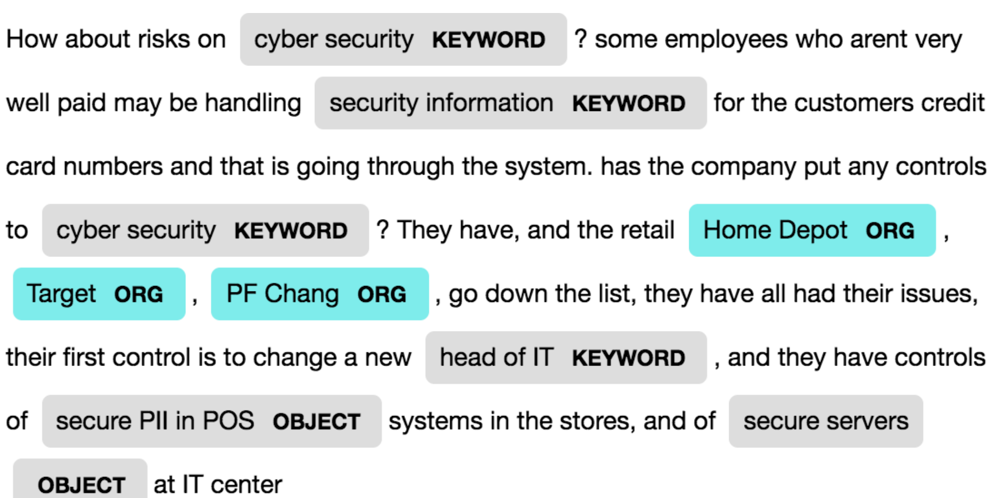


Figure 33 Example Tags of Identified Keywords and Phrases

4.7.2.2 Stage 2: Topic discovery (topic mapping)

Topic mapping has no model training process, and it is done based on the pre-defined keyword-topic table. If a sentence includes keywords that will give the sentence more than one topic, then the sentence will be tagged with all the topics, which will be used for topic linkage analysis later. If there is no keyword in a sentence, then it is removed. Then the topic(s) of each sentence is determined, Figure 34 shows some examples of labeled sentences:

How about risks on cyber security **KEYWORD** ?—topic: cyber security
some employees who arent very well paid may be handling security information **KEYWORD** for the customers credit card numbers and that is going through the system.—topic: cyber security
has the company put any controls to cyber security **KEYWORD** ?—topic: cyber security
They have, and the retail Home Depot **ORG** , Target **ORG** , PF Chang **ORG** , go down the list, they have all had their issues, their first control is to change a new head of IT **KEYWORD** , and they have controls of secure PII in POS **OBJECT** systems in the stores, and of secure servers **OBJECT** at IT center—topic: cyber security

Figure 34 Examples of Labeled Sentences

4.7.2.3 Prototype implementation for topic discovery

This section demonstrates the implementation of the topic discovery models. We start from an empty model and build it by feeding the training data. After model training, the models are tested by the testing data. The expected prototype output is that for each sentence from the testing data, the prototype can extract keywords, identify topic, and recommend its related topics. All training and testing speeches are segmented by the SBD model before fed into this module.

a. Model Training

Step 1. Initial a blank model

We first initial a blank model for English Language Processing. Multiple languages such as German, French, Spanish, Portuguese, Italian, and Dutch can be used in the model if needed. Here we give the tour example for English Language Processing only.

```
#Create a blank model
nlp = spacy.blank('en')  # create blank Language class
print("Created blank model for English Language Processing")
```

```
Created blank model for English Language Processing
```

Figure 35 Initial A Blank Model

Step 2. Load Training Data

Three training data sets stored in the training data pool (Training_data_py) are used in this implementation. Initially, we have 3 training dataset, the first dataset has 34 sentences and 44 keywords, the second dataset contains 57 sentences and 85 keywords, and the third dataset contains 11 sentences and 15 keywords. Below the first 5 sentences from the first dataset are displayed as examples.

```
#Here we load training data
from Training_data_py import *
print('The training data includes: %d data set' %len(TRAIN_DATA),'...')
for i in range(len(TRAIN_DATA)):
    print('\tDataset %d, the data includes: %d sentence and %d keywords'
          %(i,len(TRAIN_DATA[i]),sum([len(row[1]['entities']) for row in T
RAIN_DATA[i]])),'...')
print('Here are the first 5 sentences in Dataset 1')
for i in range(5):
    print('\tsentence %d:' %i,TRAIN_DATA[0][i][0])
```

```
34
The training data includes: 3 data set ...
        Dataset 0, the data includes: 34 sentence and 44 keywords ...
        Dataset 1, the data includes: 57 sentence and 85 keywords ...
        Dataset 2, the data includes: 14 sentence and 18 keywords ...
Here are the first 5 sentences in Dataset 1
        sentence 0: so are They Required to do an annual impairment test
        sentence 1: we  do end up getting an assessment of going concern
and impairment.
        sentence 2: essentially I would be surprised if there were any in
dicators that there were impairments.
        sentence 3: so has the management team been the same through all
of that
        sentence 4: or have they had much turnover?
```

Figure 36 Load Training Data

Step 3. Load Keyword-Topic List

The topic-keyword list has been initialized and is maintained by the Topic_key_list. Here the first five topics and five keywords of each topic are presented. We could append more keywords or topics by updating the Topic_key_list manually or adaptively in the online processes.

```
#here we load topic related keyword list:
from Topic_key_list import *
i=0
topic_TM=[[0 for k in range(len(topic_dic))] for kk in range(len(topic_dic
))]
for row in topic_dic:
    if i<=4:
        print('topic:', row)
        print('\t first five keyword:', topic_dic[row][:5],'......')
        i+=1
    else:
        break
```

```
load topic list:..........
topic: q_begin
         first five keyword: ['what', 'when', 'where', 'why', 'who']
......
topic: year
         first five keyword: ['2010', '2011', '2012', '2013', '2014']
......
topic: company
         first five keyword: ['AAR CORP', 'AMERICAN AIRLINES GROUP INC',
'CECO ENVIRONMENTAL CORP', 'ASA GOLD AND PRECIOUS METALS', 'AVX CORP']
......
topic: industry
         first five keyword: ['Agriculture', 'Mining', 'Utilities', 'Cons
truction', 'Manufacturing'] ......
topic: impairment
         first five keyword: ['impairment', 'intangible impairment', 'int
angible asset impairment', 'impairment of intangible assets', 'Impairment
of intangible asset'] ......
```

Figure 37 Load Keyword-Topic List

Step 4. Training Sentences Processing Pipelines

Once the training data is loaded, the keyword extraction model is trained to extract and memorize the order and combinations of labeled entities or keywords by the processing pipelines, which is used for entity prediction in the testing mode.

```
#extract keywords and text dependency from training data
ner = nlp.create_pipe('ner')
nlp.add_pipe(ner, last=True)
for row in TRAIN_DATA:
    for _, annotations in row:
        for ent in annotations.get('entities'):
            ner.add_label(ent[2])
```

Figure 38 Keyword Extraction Model Training (1)

The purpose of the training process is to build a statistical model to predict which part of a sentence is an entity or keyword, or to assign some pre-defined tags to some keywords. To train the model, we feed the training data into the following workflow and update the weight of parsed keyword combinations for optimization.

```
#train the model
import plac
n_iter=10
other_pipes = [pipe for pipe in nlp.pipe_names if pipe != 'ner']
with nlp.disable_pipes(*other_pipes):  # only train NER
    optimizer = nlp.begin_training()
    for row in TRAIN_DATA:
        for itn in range(n_iter):
            random.shuffle(row)
            losses = {}
            for text, annotations in row:
                nlp.update(
                    [text],  # batch of texts
                    [annotations],  # batch of annotations
                    drop=0.01,  # dropout - make it harder to memorise dat
a
                    sgd=optimizer,  # callable to update weights
                    losses=losses)
from IPython.display import Image
Image("training1.png",width=600)
```

Figure 39 Keyword Extraction Model Training (2)

Step 5. Topic Transition Matrix Construction

In the training process, a topic-topic linkage network is constructed for topic recommendation in the proposed online processes. If a topic “impairment” is followed by a topic discussion about “weather”, the model builds one link from “impairment” to “weather”. As more training data fed into our model, we put more weights (links) to some topic-topic pair. As a result, in the online process, if one topic is discussed, the model searches the topic-topic linkage network and provide related topic recommendation for “next topic discussion” recommendation.

To recommend "next-topic", a topic transferring matrix $\mathcal{T}$ is built. The element $\mathcal{T}_{ij}$ represents the topic transferring frequency extracted in the training records. Once topic j is detected after topic i discussion, $\mathcal{T}_{ij} = \mathcal{T}_{ij} + 1$. In online practice, if topic i is in current discussion, we would recommend the next topic according to the highest "next-topic" probability: $\mathbf{j} = \mathbf{argmax_j}\, \mathcal{T}_{ij}$.

```
#update topic_TM
for data in TRAIN_DATA:
    for i in range(len(data)-1):
        if data[i][topic_index]!=data[i+1][topic_index]:
            topic_TM[data[i][topic_index]]data[i+1][topic_index]+=1
```

Figure 40 Topic Transition Matrix

b. Model Testing

Once the model is trained, we can then put it online for testing. Here we first load a testing set including 61 sentences. Five sentences are displayed in Figure 41:

```
#once the model is trained, we can use it for testing
#here we first load the test data
from Testing_data_py import *
print('The testing data includes: %d sentences' %len(TEST_DATA),'...')
print('#Here are the first 5 sentences')
for i in range(5):
    print('sentence %d:' %i,TEST_DATA[i][0])
```

```
The testing data includes: 61 sentences ...
#Here are the first 5 sentences
sentence 0: well we got underperforming stores that may have impairment m
odels
sentence 1: there is gonna be cash flow analyses associated with those,
sentence 2: how much CAPEX are they envisioning and modeling to put into
the stores.
sentence 3: If that gonna be a significant number to improve the storefro
nts and you know bigger investment
sentence 4: that may change some of the metrics in their impairment model
s  that we probably need to factor it and not just talk to the finance te
am
```

Figure 41 Load Testing Data

The model testing is to predict the keywords of each sentence and then provide a predicted topic for the testing sentences. The output of the testing part is displayed in Figure 42.

```
#here we extract the key words, entities, and topic indicators
i=0
indicator_list=[]
for text, _ in TEST_DATA:
    doc = nlp(text)
    topic_indicator=[]
    for ent in doc.ents:
        for topic in topic_dic:
            if ent.text in topic_dic[topic]:
                topic_indicator.append(topic)
    indicator_list.append(topic_indicator)
    if i<=30:
        if len(topic_indicator)>0:
            print('sentence %d:' %i)
            print('\t keyword:',[ent.text for ent in doc.ents])
            print('\t topic indicator:',topic_indicator)
            if len(topic_indicator)>0 and topic_indicator[0] in topic_list
:
                print('\t most related topic:',topic_list[topic_TM[topic_l
ist.index(topic_indicator[0])][:].index(max(topic_TM[topic_list.index(topi
c_indicator[0])][:]))])
            print( )
        i+=1
    else: break
```

```
sentence 0:
         keyword: ['impairment']
         topic indicator: ['impairment']
         most related topic: weather

sentence 3:
         keyword: ['investment']
         topic indicator: ['impairment']
         most related topic: weather

sentence 4:
         keyword: ['impairment']
         topic indicator: ['impairment']
         most related topic: weather

sentence 25:
         keyword: ['security']
         topic indicator: ['security']
         most related topic: management
```

Figure 42 Topic Discovery Output

Here we present five testing results. As can be seen, for each sentence, the model will extract the keyword, topic, and its most related topic. In general, one sentence may cover multiple entities and keywords and one keyword may refer to multiple topics. The predicted topic is leveraged by the weight of each topic indicator.

4.7.3 Knowledge discovery - key information extraction

The knowledge discovery model is developed in the prototype to extract the most important information from a sentence or a paragraph for a determined topic. It is built using spaCy' dependency-based algorithm, which can extract relations between phrases and entities using its entity recognizer and the dependency parse, with proper "define and design" of the linguistic relationship in the dataset. With the model, we will be able to extract important knowledge from a group of sentences or paragraphs. Two things are defined for model training: what topic that we want the model to find knowledge on (also called target or context of discovers), and dependency relationships (e.g., "nsubj", "dobj", "prep with", "amod") that we define in the model to locate and extract the knowledge out of one or several complete sentences. Figure 43 shows an example paragraph in the testing data.

```
TEXTS = [
  "How about risks on cyber security",
  "some employees who aren't very well paid may be handling security
information for the customer's credit card numbers and that is going
through the system.",
  "has the company put any controls to cyber security",
  "They have, and the retail Home Depot, Target, PF Chang, go down
the list, they have all had their issues, the companies first change a new
head of IT, and they have controls of secure PII in POS systems in the
stores, and controls of secure servers at IT center"]
```

Figure 43 Example Data for Knowledge Extraction

This paragraph contains extracted brainstorming discussions on cyber security risks and related controls. To extract most important information about risks and control methods, dependencies “attr” (attributive), “dobj” (direct object: a verb before the noun phrase and the none is the target), “nsubj” (nominal subject), “pobj” (object of a preposition) and “prep” (prepositional modifier) (De Marneffe and Manning 2008) are used in our model design. Figure 44 shows the model processing and the output of knowledge discovery.

```
def extract_risk_control_relations(doc):
    # merge entities and noun chunks into one token
    spans = list(doc.ents) + list(doc.noun_chunks)
    for span in spans:
        span.merge()
    relations = []
    for control in filter(lambda w: w.ent_type_ == 'keyword', doc):
        if control.dep_ in ('attr', 'dobj'):
            subject = [w for w in control.head.lefts if w.dep_ == 'nsubj']
            if subject:
                subject = subject[0]
                relations.append((subject,control.head,control))
        elif control.dep_ == 'pobj' and control.head.dep_ == 'prep':
            relations.append((control.head.head,control.head,control))
    return relations
```

```
from spacy import displacy
from spacy.tokens import Doc
for text1 in TEXTS:
    doc1=nlp(text1)
    i=0
    relations = extract_risk_control_relations(doc1)
    for r1,r2,r3 in relations:
        print('{:<10}\t{}\t{}'.format(r1.text, r2.text, r3.text))
```

```
risks           on      cyber security
some employees  handling        security information
put             to      cyber security
the companies   change  head of IT
controls        of      secure PII in POS
controls        of      secure servers
```

Figure 44 Example Output of Knowledge Discovery

4.7.4 Knowledge adaptive learning and Question Answering

Historical knowledge should be stored in the knowledge base of the proposed system. New knowledge in future brainstorming conversation cases can be automatically collected and stored under the correct “topic” of the knowledge base. In this proposed prototype, after the testing process, there will be new knowledge from the testing data being discovered. For example, topic “impairment” was discussed in our testing set and the sentences about this topic will be stored into the knowledge base. Part of the new knowledge obtained from the testing set is shown in Figure 45 below.

```
current_topic=indicator_list[0][0]
current_sentence=0
current_same_topic_sentence={}
current_same_topic_sentence[current_topic]=TEST_DATA[0][0]
for i in range(len(indicator_list)-1):
    if len(indicator_list[i+1])>0:
        if current_topic in indicator_list[i+1]:
            current_same_topic_sentence[current_topic]+=TEST_DATA[i+1][0]
        else:
            current_topic=indicator_list[i+1][0]
            current_same_topic_sentence[current_topic]=[TEST_DATA[i+1][0]]
    else:
        current_same_topic_sentence[current_topic]+=TEST_DATA[i+1][0]
print('the achieved knowledge is about topic %s'%current_same_topic_senten
ce.keys())
for key in current_same_topic_sentence:
    print('topic', key, 'contains', current_same_topic_sentence)
```

```
the achieved knowledge is about topic dict_keys(['impairment', 'securit
y'])
topic impairment contains {'impairment': 'well we got underperforming sto
res that may have impairment modelsthere is gonna be cash flow analyses a
ssociated with those,how much CAPEX are they envisioning and modeling to
put into the stores.If that gonna be a significant number to improve the
storefronts and you know bigger investment  that may change some of the m
etrics in their impairment models  that we probably need to factor it and
not just talk to the finance team and talk to those that are making those
operational decisions  to make sure the finances is capturing those types
of projections appropriately.What about geographic risk? We talk about th
em being in the Southeast, Midwest, ok and then you said they tried to ex
pand a little bit in the west and didnt go as well.I think mostly is Midw
est and Southeastand they have some up in the Northeast as well you know
```

Figure 45 Knowledge Adaptive Learning

From the testing set, this model identified 10 sentences that were tagged with “impairment” and saved them for future analysis. If a future user wants to know if the topic "impairment" was discussed before and what knowledge was collected about it, they can raise questions to the prototype through the QA module.

QA module allows a future user to input queries. In our example, if we ask the prototype "Do they have any impairment risks before?" after the testing process, the system will parse the keywords and recognize the query topic "impairment" and then search for related knowledge. If the topic was discussed before, the module would output the past discussions, which were collected from our previous online testing. Only information sentences that contain keywords in the keyword-topic table will be considered as useful information and be used for further knowledge extraction, while other sentences such as agreement (e.g., "yeah, I agree") or greetings will be removed. Figure 46 shows part of the answer generated through the QA module for this question.

```
def topic_query():
    question=input("Query:")
    doc=nlp(str(question))
    topic_indicator=[]
    for ent in doc.ents:
        for topic in topic_dic:
            if ent.text in topic_dic[topic]:
                topic_indicator.append(topic)
    for topic in topic_indicator:
        if topic in current_same_topic_sentence:
            print('YES the topic %s was discussed before' %topic)
            print('Its related topic is %s' %topic_list[topic_TM[topic_lis
t.index(topic_indicator[0])][:].index(max(topic_TM[topic_list.index(topic_
indicator[0])][:]))])
            print('here is the past discussion:')
            print(current_same_topic_sentence[topic])
        else:
            print('This topic was not discussed before')
topic_query()
```

```
Query:do they have any impairment risk before
YES the topic impairment was discussed before
Its related topic is weather
here is the past discussion:
well we got underperforming stores that may have impairment modelsthere i
s gonna be cash flow analyses associated with those,how much CAPEX are th
ey envisioning and modeling to put into the stores.If that gonna be a sig
nificant number to improve the storefronts and you know bigger investment
  that may change some of the metrics in their impairment models  that we
probably need to factor it and not just talk to the finance team and talk
```

Figure 46 Question Answering Example

4.8 Application of the cognitive assistant to other audit phases

Besides the audit plan brainstorming process, the proposed audit cognitive assistant can be extended and applied to other audit phrases. As discussed in Chapter 2, there are four main features/advantages of a cognitive assistant that make it potentially an excellent tool for the audit field: information retravel support, recommendation support, adaptive learning capability and service delegation. The major phases of an audit include client acceptance, preliminary engagement activities, audit plan, internal control audit, business processes and accounts audit, audit completion (Messier et al. 2016). This section will introduce how to extend the proposed cognitive assistant design to each of these audit phases.

The following questions will be discussed for each audit phrase:

1. Scenarios to use the proposed cognitive assistant and users
2. Scenarios to apply the proposed audit plan knowledge discovery system (APKDS) framework
3. Design of the important modules for the phrase: knowledge base (QA categories), recommendations, and related apps

Table 14 shows a summary of different audit stages and how the proposed cognitive assistant can be used to provide decision supports to each of the stages.

	client acceptance	preliminary engagement activities	audit plan	internal control audit	business processes and accounts audit	audit completion
Decision support in the cognitive assistant						
- Information Retrieval	√	√	√	√	√	√
- Recommendations	√	√	√	√	√	√
-App connections	√	√	√	√	√	√
Collected Knowledge						
-NLP conversation analysis framework	√		√	√		√

Table 14 Application of The Proposed Tool to Different Audit Stages

4.8.1 Client acceptance

In the beginning of an audit engagement, the audit firm needs to make the decision on whether or not to accept new clients and to retain current clients (Messier et al. 2016; Hsieh and Lin 2015). The knowledge that auditors gather during the acceptance/continuance process provides valuable information for auditors to understand the client and its environment. Auditors can interact with the system for decision aids either when they work individually, or when there is a discuss meeting among partners or between partner(s) and the client.

In the client acceptance process, auditors communicate with different parties and consider possible unusual business or audit risks (Messier et al. 2016). The NLP-based audit conversation analysis framework can help collect and extract important knowledge from those communications. Collected knowledge provides expertise and experience of senior auditors in assessing a client. The related communication scenarios include:

- teleconference or meeting among auditors (among partners or partner and auditors)
- inquiries made by successor auditor to the predecessor auditor
- discussions with third parties (bankers and attorneys of client, credit agencies, legal counsel and industry peers, and other members of the business community)

To provide information support for the acceptance/continuance process, related knowledge base of the proposed cognitive assistant needs to be built with two main types of information: information extracted from available files (entity documents and reports) and collected knowledge from inquiries and meetings. If useful information from observations, inquires and meeting discussions can be effectively collected and documented, they could then be used as available files in the knowledge base for future engagement cases.

Prior studies (Messier et al. 2016) suggest that auditors should consider management, financial health, company background, audit risk, litigation risk, expertise of engagement personnel, audit fee etc. when determining if they should accept the engagement. When making a client continuance (whether to continue with current clients) decision, auditors should also consider significant events and potential conflicts or disputes (Hollingsworth 2012; Hsieh and Lin 2015).

Based on the risk areas summarized in prior study on client acceptance/continuance process (Hollingsworth 2012; Messier et al. 2016), Table 15 shows proposed question categories that should be built for the QA module in the audit cognitive assistant. Related knowledge can be collected and then organized according to these question categories in order to support information retrieval requests during this audit process.

Decision Support by cognitive assistant	**Detailed design**		**Technical support and knowledge sources**
Information Retrieval	**Question categories**	**Sub-question categories**	
	Management	Management integrity, Management tone toward financial reporting, management competence, turnover of key personnel;	Available files: entity documents and reports Collected knowledge: Documented observations and inquires; converted inquires and meeting discussions
	Financial health	Financial distress, financial leverage, profitability of company, company cash flow, stock volatility	
	Company characteristics	Complexity of company, business changes, Accounting & Finance Dept personnel competence, internal controls, litigation risk/ prior restatements, significant related party transactions, strain the client puts on the audit firm's staff, reasons for change of auditors	
	Governance	Audit committee competence, Audit committee financial expertise	

	Auditor characteristics	Availability of appropriate personnel, sufficient expertise of personnel, profitability of engagement	
	Client Continuance	Significant event; conflicts over accounting and auditing issues; disputes over fees	
Recommendation	Sub question categories; Updated related topics discovered from adaptive learning module		Pre-defined recommendation table; Adaptive learning module
Applications	Audit programs/tools on internal controls		Service Delegation module

Table 15 Proposed Question Categories for Client Acceptance

Knowledge sources:

Table 16 proposes detailed knowledge sources for the proposed question categories in Table 15. Listed available files are unstructured data that can be stored in the knowledge base for both IR-based QA and knowledge-based (cognitive) QA.

Knowledge type	**Details**	**Format to save in the QA system**
Available files	- financial information: annual reports, interim financial statements, income tax returns, etc. - regulations, audit standards - received inquiry from third parties - notes from meetings discussion - memo or completed entity acceptance questionnaire or checklist - documented inquiries from predecessor auditor - copies of working papers from predecessor auditor - risk rating from audit programs	IR-based QA system and knowledge QA system

Collected knowledge	- converted knowledge from conversational teleconference, meetings, inquiries	Knowledge QA system

Table 16 Knowledge Sources for Client Acceptance

4.8.2 Preliminary Engagement Activities

In the preliminary engagement activities, there are generally three main tasks: determine the audit engagement team requirements, ensure audit team's independence, understand services to be performed and the other terms of the engagement. In this audit phase, since related auditors mainly complete the activities individually with the help of developed audit programs, they can interact with the cognitive assistant for information retrieval support when they review prior documents and make judgements individually. The NLP-based audit conversation analysis framework does not need to be used at this stage. Based on Messier et al. (2016) study on risk areas in the preliminary engagement activities stage, Table 17 summarizes the proposed question categories and knowledge sources for preliminary engagement activities.

Decision Support by cognitive assistant	**Detailed design**		**Technical support and knowledge sources**
Information retrieval	**Question categories**	**Sub-question categories**	
	Engagement team requirements	Engagement size and complexity, level of risk, special expertise of the audit team[13], time budget, personnel availability	Available files: - annual independence questionnaire

[13] if a specialized industry (such as banking, insurance) is involved, or if the client has sophisticated IT or holds financial instruments, requisite expertise of audit team members must be ensured

	Audit firm independence	Profession's ethical requirements, auditor independence, unpaid client fees, consulting services	- annual independence reports - Engagement Letter
	Consulting services		
Recommendation	Sub-question categories; Industry specific recommendations (e.g., special expertise of the engagement team for a specialized industry) Updated related topics discovered from adaptive learning module		Pre-defined recommendation table; Adaptive learning module
Applications	Audit programs/tools: audit firm's independent database (fill new form) Memo/checklist; Auditor personnel management grogram (check availability)		Service Delegation module

Table 17 Proposed Question Categories and Knowledge Sources for Preliminary Engagement Activities

4.8.3 Internal Control Audit

Understanding of internal control can help auditors identify and assess risk and find areas where financial statements may be misstated. The cognitive assistant could be used by the engagement team when they try to obtain a better understanding about client's internal control system and when they conduct inquiries or meetings with management. The NLP-based audit conversation analysis framework can be applied if auditors communicate with appropriate management, supervisory, and staff personnel to collect internal control information through spoken dialogues. Based on the risk factors summarized in prior study (Messier et al. 2016) on internal controls, Table 18 summarizes the proposed question categories for internal control audit and Table 19 proposes related knowledge sources.

Decision Support by cognitive assistant	**Detailed design**		**Technical support and knowledge sources**
Information retrieval	**Question categories**	**Sub-question categories**	Available files; Collected knowledge
	Entity’s operations and systems	Manual controls and automated controls; whether an IT specialist is needed	
	Control Environment	Management’s and the board of directors’ attitudes, awareness, and actions	
	Control Activities		
	Information and Communication		
	Monitoring Activities		
Recommendation	Sub-question categories; Updated related topics from adaptive learning module		Pre-defined recommendation table; Adaptive learning module
Applications	Audit programs/tools: internal control questionnaire within a firm’s audit software		Service Delegation module

Table 18 Proposed Question Categories for Internal Control Audit

Knowledge source:

Knowledge type	**Details**	**Format to save in the QA system**
Available files	- written communication records on control deficiencies with management (template, notes) - Documented understanding of entity’s internal controls in recurring engagement case	

	- entity's procedures manuals (entity's policies and procedures such as accounting systems and related control activities) - Organizational Charts - Internal control questionnaires - Flowcharts of the entity's accounting system	IR-based QA system and knowledge QA system
Collected knowledge	- extracted discussion (control deficiencies, fraud or illegal acts) from meetings with management and the audit committee	Knowledge QA system

Table 19 Proposed Knowledge Sources for Internal Control Audit

4.8.4 Business Processes and Related Accounts Audit

In this audit phase, audit procedures will be applied to the accounts so that auditors obtain audit evidence about management's assertions relating to each account and lower the risk of undetected material misstatement.

Audit firms usually spent a great amount of time on the financial statement audit and the audit of internal control over Financial Reporting. In this process, auditors can use the proposed cognitive assistant to look up for information needed to make judgments. The NLP-based conversation analysis framework may not be used in this phase if there are no multi-party meetings. Based on the common risk areas in business processes and related accounts that were summarized in prior study (Messier et al. 2016), Table 20 below proposes how to develop question categories for important business processes and significant accounts, and what knowledge sources should be considered to build QA system knowledge base. The knowledge sources come from important available documents and records.

Decision Support by cognitive assistant	**Detailed design**	**Technical support and knowledge sources**

Information retrieval	**Question categories**	**Sub-question categories**	
	Revenue process	· Industry-related factors. · The complexity and contentiousness of revenue recognition issues. · The difficulty of auditing transactions and account balances. · Misstatements detected in prior audits.	Customer sales order, Credit approval form, Open-order report, Shipping document, Sales invoice, Sales journal, Customer statement, Accounts receivable subsidiary ledger, Aged trial balance of accounts receivable, Remittance advice, Cash receipts journal, Credit memorandum, Write-off authorization
	Purchasing Process	Industry-Related Factors; Misstatements Detected in Prior Audits	Purchase requisition, Voucher register/purchases journal, Purchase order, Accounts payable subsidiary ledger, Receiving report, Vendor statement, Vendor invoice, Check/EFT, Voucher Cash, disbursements journal/check register
	Payroll Application	Turnover; The presence of labor contracts and legislation (such as the Occupational Safety and Health Act);	Personnel records, including wage-rate or salary authorizations W-4 and other deduction authorization forms Time card/time sheet Payroll check/direct deposit records Payroll register Payroll master file Payroll master file changes report

			Periodic payroll reports Various tax reports and forms
	Inventory Management Process	Industry-related Factors; Operating and engagement characteristics	Production schedule, Production data information, Receiving report, Cost accumulation and variance report, Materials requisition, Inventory status report, Inventory master file, Shipping order
	Property, plant, and equipment	Complex accounting issues. · Difficult-to-audit transactions. · Misstatements detected in prior audits.	
	Long-Term Liabilities, Stockholders' Equity, and Income Statement Accounts		Financial statements
	Cash and Investments		· bank reconciliation working paper · Standard Bank Confirmation Form · A cutoff bank statement.
Recommendation	Sub-question categories; Updated related topics from adaptive learning module		Pre-defined recommendation table; Adaptive learning module
Applications	Client ERP system with related business process sub-system		Service Delegation module

Table 20 Proposed Question Categories and Knowledge Sources for Business Processes and Related Accounts Audit

4.8.5 Complete the Audit Engagement

After auditors finishing gathering reliable evidence and auditing financial statement accounts and related controls of business processes, they will summarize and evaluate the evidence. In this phase, auditors use sufficient appropriate evidence to reach a conclusion on the fairness of the financial statements (Messier et al. 2016). In addition, auditors apply audit procedures to identify undisclosed contingent liabilities and search for possible subsequent events that occur after the balance sheet date but before the date of the audit report. Auditors can interact with the cognitive assistant for decision aids either when they review and examine documentations individually, or they have two-way dialogues with management and legal counsel. The NLP-based audit conversation analysis framework can help collect and extract important knowledge from those discussions.

Based on the risk areas identified by prior studies (Messier et al. 2016), Table 21 proposes how to develop question categories for identifying contingent liabilities and subsequent events. Table 22 proposes the detailed files and scenarios where the NLP-based framework can be used to collect knowledge.

Decision Support by cognitive assistant	**Detailed design**	**Technical support and knowledge sources**
Information Retrieval	**Question categories**	Available files; Collected knowledge
	Contingent liabilities	

	Subsequent events for financial statements audit	
	Subsequent events for internal control audit over financial reporting	
	Going concern consideration	

Table 21 Proposed Question Categories for Complete the Audit

Knowledge source:

Knowledge type	**Details**	**Format to save in the QA system**
Available files	- working papers - financial statements - minutes of meetings of the board of directors, committees of the board, and stockholders - contracts, leases, loan agreements, and correspondence from government agencies - tax returns, IRS reports - guarantees and letters of credit from financial institutions or lending agencies - correspondence and invoices from attorneys for pending or threatened lawsuits from client - a written representation from management with all litigation, claims, and assessments disclosed - internal audit reports - Independent auditor reports of significant deficiencies or material weaknesses - Regulatory agency reports of internal control - management representation letter - management letter (include recommendations to client)	IR-based QA system and knowledge QA system
Collected knowledge	- converted knowledge from spoken discussions with management - converted knowledge from inquires of legal counsel about litigation, claims, and assessments against the client - converted knowledge from communications with those charged with governance and management	Knowledge QA system

Table 22 Proposed Knowledge Sources for Complete the Audit

4.9 Summary

This paper proposes the development of a prototype for the NLP-based audit plan knowledge discovery system and illustrates the development of important modules. This proposed development shows the feasibility and functionality of the audit conversation analysis framework proposed in Chapter 3. Detailed design of sentence-keyword table and keyword-topic table for building this audit domain specific tool is also provided.

With the help of Natural Language Processing (NLP) techniques, this proposed development of the prototype demonstrate that the proposed NLP-based audit plan knowledge discovery system could potentially be built and be used to collect knowledge and experience from senior auditors and specialists, transfer this special type of knowledge into standardized database format. Moreover, the engagement teams can ask questions through a Question-Answering module to communicate with the system, and they will receive answers and recommendations on risk related topics as decision aids. This prototype demonstration also shows that the proposed system will continuously collect and update its knowledge base.

CHAPTER 5: CONCLUSION AND FUTURE WORK

In this chapter, we conclude the dissertation first with an overall review and then discuss about future research directions.

5.1 Review of Major Results

Auditors' discussions during audit brainstorming meetings involve various topics on risks, which provide numerous valuable knowledge on how auditors identify inherent and fraud risks and how they make audit decisions. This knowledge is difficult to collect and analyze, but once collected, it can be integrated into computer assisted audit tools and techniques (CAATTs) and provide decision support to engagement teams in future audit engagement cases.

This dissertation attempts to contribute to the auditing literature and practice by introducing the emerging AI-based cognitive assistant technology into auditing domain and providing a vision of future audit. An interactive intelligent cognitive system framework has been proposed in this thesis, which aims at helping auditors identify and assess risks during audit plan brainstorming sessions and make subsequent decisions. The following summarizes the three essays.

The first essay proposes an audit domain cognitive assistant framework that could provide interactive decision aids to auditors in the audit brainstorming meetings. The proposed audit cognitive assistant framework has several important modules such as natural language processing module, question answering system, recommendation system and service delegation system. Like other existing commercial cognitive assistants, this audit cognitive assistant is supposed to be able to understand users' questions and commands in natural language and generate answers and recommendations for them. The proposed audit cognitive assistant provides a new methodology for audit knowledge organization and knowledge collection. Since a cognitive assistant has the adaptive learning capability, this proposed audit tool can potentially create a knowledge base that stores many senior auditors' professional knowledge and experience on audit risk assessment and audit decision making, which could also be used as valuable knowledge sources to support other computer-assisted audit tools and techniques (CAATTs) in an audit firm.

The second essay proposes a framework of an NLP-based audit plan knowledge discovery system (APKDS). The framework is designed to collect and analyze audit brainstorming discussion contents automatically and continuously. Extracted and summarized discussion contents will then be converted into standardized data in the knowledge base for future information retrieval support and decision-making support in various audit tools, including the audit cognitive assistant proposed in essay one. Although there have been many NLP-related studies in different fields, to the best of our knowledge, this is the first NLP-based system framework proposed for audit domain conversation analysis. It aims at programing machine learning algorithms to process large amounts of natural language data in auditing documents.

The third essay proposes the development of an APKDS prototype. We demonstrate the development of proposed modules using the NLP software spaCy and with extracted audit risk assessment knowledge from four experimental brainstorming cases. This paper shows the potential feasibility and functionality of the proposed NLP conversation analysis framework. In addition, we provide a discussion on the potential application of the proposed cognitive assistant to other important audit phases, such as client acceptance, preliminary engagement activities, internal control audit, business processes and accounts audit, and audit completion.

5.2 Future Work

Because of information explosion and technology innovations, future audit will require information analysis and knowledge extraction from large data populations and from various information sources. It will be necessary to have intelligent and effective audit tools that can provide auditors information retrieval and decision-making aids in time. In general, my current and future research focus will still be on audit decision support system development and audit knowledge discovery and management. I would like to contribute to the systematic research on audit data analytics and audit intelligence.

Yet, this study has some limitations and we will work on them in our future work. First of all, this current design of the prototype was trained and tested only with a small amount of sample data, which should be extended with larger data size for model improvement in the future work. Further, since system evaluation is a crucial step in the design science research, more detailed system evaluation process should be designed based the proposed evaluation method and should be realized to test the proposed system as well.

According to recent studies (Ebling 2016; Hu et al 2017), the improvement of cognitive assistant technology will make it more talented in understanding natural language, understanding context of request, and even understanding nonverbal language such as emotions from voice tone and facial expressions and gestures. If these improvements could be realized in the future, the proposed audit cognitive assistant can be improved in design and could better serve for audit plan risk assessment as it can better understand the engagement team during the brainstorming group discussions based on auditors' nonverbal language.

BIBLIOGRAPHY

Abreu, R., David, F., & Segura, L. (2014). The management of accounting and taxation knowledge in Portugal. In Garlatti, A. & Massaro, M. (Eds.). Proceedings of the 15th European Conference on Knowledge Management (Vol. 1, pp. 8–14). Academic Conferences and Publishing International.

AICPA (2002). AU Section 316: Consideration of Fraud in a Financial Statement Audit. https://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00316.pdf

AICPA (2012). The Risk Assessment Auditing Standards: How to Efficiently and Effectively Comply on Smaller and Less Complex Audit Engagements. *Whitepaper.*

Asare, S. K., & Wright, A. M. (2004). The effectiveness of alternative risk assessment and program planning tools in a fraud setting. *Contemporary Accounting Research*, *21*(2), 325-352.

Banerjee, S., Rose, C., & Rudnicky, A. I. (2005, September). The necessity of a meeting recording and playback system, and the benefit of topic–level annotations to meeting browsing. In *IFIP Conference on Human-Computer Interaction* (pp. 643-656). Springer, Berlin, Heidelberg.

Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, *60*(6), 1371-1391.

Beasley, M. S., & Jenkins, J. G. (2003). A primer for brainstorming fraud risks. *JOURNAL OF ACCOUNTANCY-NEW YORK-*, *196*(6), 32-40.

Bellegarda, J. R. (2013). Large-scale personal assistant technology deployment: the Siri experience. In INTERSPEECH (pp. 2029-2033).

Bellovary, J. L., & Johnstone, K. M. (2007). Descriptive evidence from audit practice on SAS No. 99 brainstorming activities. *Current Issues in Auditing*, *1*(1), A1-A11.

Bohnet, B. (2010, August). Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd international conference on computational linguistics*(pp. 89-97). Association for Computational Linguistics.

Bonner, S. E. (2008). Judgment and Decision Making in Accounting. Upper Saddle River, NJ: Pearson/Prentice Hall.

Bost, X., Senay, G., El-Bèze, M., & De Mori, R. (2015). Multiple topic identification in human/human conversations. *Computer Speech & Language*, *34*(1), 18-42.

Bouthillier, F. & Shearer, K. (2002). Understanding knowledge management and information management: The need for an empirical perspective. Information research, 8(1), 8–1.

Braun, D., Hernandez-Mendez, A., Matthes, F., & Langen, M. (2017). Evaluating natural language understanding services for conversational question answering systems. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 174-185).

Britta O'boyle (2015).What is Siri? Apple's personal voice assistant explained. http://www.pocket-lint.com/news/112346-what-is-siri-apple-s-personal-voice-assistant-explained

Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Datas impact on audit judgment and decision making and future research directions. Accounting Horizons, 29(2), 451–468. doi:10.2308/ acch-51023

Burger, S., MacLaren, V., & Yu, H. (2002). The ISL meeting corpus: The impact of meeting type on speech style. In *Seventh International Conference on Spoken Language Processing*.

Cambria, E., & White, B. (2014). Jumping NLP curves: a review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, *9*(2), 48-57.

Canbek, N. G., & Mutlu, M. E. (2016). On the track of Artificial Intelligence: Learning with intelligent personal assistants. *Journal of Human Sciences*, *13*(1), 592-601.

Carpenter, T. D. (2007). Audit team brainstorming, fraud risk identification, and fraud risk assessment: Implications of SAS No. 99. *The Accounting Review*, *82*(5), 1119-1140.

Chen, A. N., Goes, P. B., & Marsden, J. R. (2002). A query-driven approach to the design and management of flexible database systems. *Journal of Management Information Systems*, *19*(3), 121-154.

Chen, D., & Manning, C. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 740-750).

Chen, G., Parada, C., & Sainath, T. N. (2015, April). Query-by-example keyword spotting using long short-term memory networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 5236-5240). IEEE.

Chen, K. Y., Liu, S. H., Chen, B., & Wang, H. M. (2018). An Information Distillation Framework for Extractive Summarization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *26*(1), 161-170.

Choi, J. D., & McCallum, A. (2013). Transition-based dependency parsing with selectional branching. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 1052-1062).

Choi, J. D., Tetreault, J., & Stent, A. (2015). It depends: Dependency parser comparison using a web-based evaluation tool. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (Vol. 1, pp. 387-396).

Chung, H., Song, Y. I., Han, K. S., Yoon, D. S., Lee, J. Y., Rim, H. C., & Kim, S. H. (2004, July). A practical QA system in restricted domains. In *Proceedings of the Workshop Question Answering in Restricted Domains, within ACL*.

Daelemans, W. (2013, March). Explanation in computational stylometry. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 451-462). Springer, Berlin, Heidelberg.

Dai, J., & Li, Q. (2016). Designing Audit Apps for Armchair Auditors to Analyze Government Procurement Contracts. *Journal of Emerging Technologies in Accounting*, *13*(2), 71-88.

Dalva, D., Guz, U., & Gurkan, H. (2017). Effective semi-supervised learning strategies for automatic sentence segmentation. *Pattern Recognition Letters*.

Davenport, Thomas (1997), “Knowledge Management at Ernst & Young”, Available http://www.mccombs.utexas.edu/kman/E&Y.htm

De Marneffe, M. C., & Manning, C. D. (2008). *Stanford typed dependencies manual* (pp. 338-345). Technical report, Stanford University.

De Mori, R., Bechet, F., Hakkani-Tur, D., McTear, M., Riccardi, G., & Tur, G. (2008). Spoken language understanding. *IEEE Signal Processing Magazine*, *25*(3).

Dowling, C., & Leech, S. (2007). Audit support systems and decision aids: Current practice and opportunities for future research. *International Journal of Accounting Information Systems*, *8*(2), 92-116.

Dowling, C., Leech, S. A., & Moroney, R. (2008). Audit support system design and the declarative knowledge of long-term users. *Journal of Emerging Technologies in Accounting*, *5*(1), 99-108.

Ebling, M. R. (2016). Can cognitive assistants disappear?. *IEEE Pervasive Computing*, *15*(3), 4-6.

Ericsson, K. A. (6). 8: Simon, HA (1984). Protocol analysis: Verbal reports as data.

Favre, B., Grishman, R., Hillard, D., Ji, H., Hakkani-Tur, D., & Ostendorf, M. (2008, March). Punctuating speech for information extraction. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 5013-5016). IEEE.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., ... & Schlaefer, N. (2010). Building Watson: An overview of the DeepQA project. *AI magazine*, *31*(3), 59-79.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137-144.

Garrido, P., Martinez, F., & Guetl, C. (2010). Adding semantic web knowledge to intelligent personal assistant agents. In *Proceedings of the ISWC 2010 Workshops* (pp. 1-12). Philippe Cudre-Mauroux.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992, March). SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*(Vol. 1, pp. 517-520). IEEE.

Goffman, Neal, (2016). The Rise of the Cognitive Assistant. https://www.ibm.com/blogs/insights-on-business/government/rise-cognitive-assistant/

Greenman, C. (2017). Exploring the Impact of Artificial Intelligence on the Accounting Profession. *Journal of Research in Business, Economics and Management*, *8*(3), 1451-1454.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, *37*(2).

Hauswald, J., Laurenzano, M. A., Zhang, Y., Li, C., Rovinski, A., Khurana, A & Mars, J. (2015). Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers. In *ACM SIGPLAN Notices* (Vol. 50, No. 4, pp. 223-238). ACM.

Hazen, T. J. (2011). Topic identification. *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, 319-356.

Hearst, M. A. (2011). 'Natural' search user interfaces. *Communications of the ACM*, *54*(11), 60-67.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly*, *28*(1), 6.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261-266.

Hoffman, V. B., & Zimbelman, M. F. (2009). Do strategic reasoning and brainstorming help auditors change their standard audit procedures in response to fraud risk?. *The Accounting Review*, *84*(3), 811-837.

Hollingsworth, C. (2012). Risk Management in the Post-SOX Era. *International Journal of Auditing*, *16*(1), 35-53.

Honnibal, M., & Montani, I. (2017). spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*.

Hsieh, Y. T., & Lin, C. J. (2015). Audit firms' client acceptance decisions: Does partner-level industry expertise matter?. *Auditing: A Journal of Practice & Theory*, *35*(2), 97-120.

Hsueh, P. Y., & Moore, J. D. (2007). What decisions have you made?: Automatic decision detection in meeting conversations. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference* (pp. 25-32).

Hu, T. Y., Raman, C., Maza, S. M., Gui, L., Baltrusaitis, T., Frederking, R. E., ... & Eskenazi, M. (2017, February). Integrating Verbal and Nonvebval Input into a Dynamic Response Spoken Dialogue System. In *AAAI* (pp. 5091-5092).

Huang KW, Li ZL (2011) A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. ACM Trans. Management Inform. Systems 2(3):1–19.

Huggins-Daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M., & Rudnicky, A. I. (2006, May). Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* (Vol. 1, pp. I-I). IEEE.

Hunton, J. E., & Gold, A. (2010). A Field Experiment Comparing the Outcomes of Three Fraud Brainstorming Procedures: Nominal Group, Round Robin, and Open Discussion (Retracted). *The Accounting Review*, *85*(3), 911-935.

Hutchinson, B., Zhang, B., & Ostendorf, M. (2010, March). Unsupervised broadcast conversation speaker role labeling. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (pp. 5322-5325). IEEE.

IBM (2011). Watson - A System Designed for Answers. *IBM Systems and Technology, An White Paper*

IBM (2016). KPMG Announces Agreement with IBM Watson to Help Deliver Cognitive-Powered Insights. https://www-03.ibm.com/press/uk/en/pressrelease/49309.wss

Institute of Chartered Accountants of Scotland (ICAS). (2012). A professional judgement framework for financial reporting: An international guide for preparers, auditors, regulators and standard setters. Edinburgh: Technical Policy & Services Board of The ICAS.

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), 1-20.

Janin, A., Ang, J., Bhagat, S., Dhillon, R., Edwards, J., Macias-Guarasa, J., ... & Wooters, C. (2004, May). The ICSI meeting project: Resources and research. In *Proceedings of the 2004 ICASSP NIST Meeting Recognition Workshop*.

Janvrin, D., Lowe, D. J., & Bierstaker, J. (2008). Auditor acceptance of computer-assisted audit techniques. *Iowa State University, Arizona State University and Villanova University*, *4*.

Klersey, G. F., & Mock, T. J. (1989). Verbal protocol research in auditing. *Accounting, Organizations and Society*, *14*(1-2), 133-151.

Kokina, J., & Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing. *Journal of Emerging Technologies in Accounting*.

Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency parsing. *Synthesis Lectures on Human Language Technologies*, *1*(1), 1-127.

Landis, M., Jerris, S. I., & Braswell, M. (2008). Better brainstorming. *Journal of Accountancy*, *206*(4), 70.

Lee, D. (2016). KPMG Recruits IBM Watson for Cognitive Tech Audits, Insights. https://www.accountingtoday.com/news/kpmg-recruits-ibm-watson-for-cognitive-tech-audits-insights

Lei, T., Xin, Y., Zhang, Y., Barzilay, R., & Jaakkola, T. (2014). Low-rank tensors for scoring dependency structures. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 1381-1391).

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2, pp. 302-308).

Levy, Steven (2016). An exclusive inside look at how artificial intelligence and machine learning work at Apple. https://backchannel.com/an-exclusive-look-at-how-ai-and-machine-learning-work-at-apple-8dbfb131932b

Lisowska, A. (2003). Multimodal interface design for the multimodal meeting domain: Preliminary indications from a query analysis study. *Report IM2. MDM-11*.

Liu, Y., & Xie, S. (2008, March). Impact of automatic sentence segmentation on meeting summarization. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 5009-5012). IEEE.

Luger, G. F. (2005). *Artificial intelligence: structures and strategies for complex problem solving*. Pearson education.

Lynch, A. L., Murthy, U. S., & Engle, T. J. (2009). Fraud brainstorming using computer-mediated communication: The effects of brainstorming technique and facilitation. *The Accounting Review*, *84*(4), 1209-1232.

Mahzan, N., & Lymer, A. (2014). Examining the adoption of computer-assisted audit tools and techniques: Cases of generalized audit software use by internal auditors. *Managerial Auditing Journal*, *29*(4), 327-349.

Martins, A., Almeida, M., & Smith, N. A. (2013). Turning on the turbo: Fast third-order non-projective turbo parsers. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2, pp. 617-622).

Matusov, E., Hillard, D., Magimai-Doss, M., Hakkani-Tür, D., Ostendorf, M., & Ney, H. (2007). Improving speech translation with automatic boundary prediction. In *Eighth Annual Conference of the International Speech Communication Association*.

Mehrez, T., Abdelkawy, A., Heikal, Y., Lange, P., Nabil, H., & Suendermann-Oeft, D. (2013). Who discovered the electron neutrino? A telephony-based distributed open-source standard-compliant spoken dialog system for question answering. *Proc. of the GSCL, Darmstadt, Germany*, *80*.

Messier, W. F., Glover, S. M., & Prawitt, D. F. (2016). *Auditing and assurance services: A Systematic Approach 10th Edition.* McGraw-Hill.

Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D. L., ... & Tambe, M. (2007). An intelligent personal assistant for task and time management. *AI Magazine*, *28*(2), 47.

National Research Council. (2007). *Human-system integration in the system development process: A new look*. National Academies Press.

Nelson, M. (2009). A model and literature review of professional skepticism in auditing. Auditing: A Journal of Practice & Theory, 28(No. 2), 1e34. November 2009.

Nguyen, L., & Kohda, Y. (2017). Toward a Knowledge Management Framework for Auditing Processes. *International Journal of Knowledge and Systems Science (IJKSS)*, *8*(3), 45-67.

Nguyen, L., Umemoto, K., Kohda, Y., & Blake, J. (2015). Knowledge Management in Auditing: A Case Study in Vietnam". In M. Massaro, & A. Garlatti (Eds.) Proceedings of the 16th European Conference on Knowledge Management (ECKM), Udine, Italy (pp. 571–577).

Nigrini, M. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection* (Vol. 586). John Wiley & Sons.

PCAOB (2010). AS 2100 Identifying and Assessing Risks of Material Misstatement. https://pcaobus.org/Standards/Auditing/Pages/AS2110.aspx

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012, May). Design science research evaluation. In *International Conference on Design Science Research in Information Systems* (pp. 398-410). Springer, Berlin, Heidelberg.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... & Silovsky, J. (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding* (No. EPFL-CONF-192584). IEEE Signal Processing Society.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A taxonomy of evaluation methods for information systems artifacts. *Journal of Management Information Systems*, *32*(3), 229-267.

Ramos, M. (2003). Auditors' responsibility for fraud detection. *Journal of Accountancy*, *195*(1), 28-36.

Raphael, J. (2017). Rethinking the Audit: Innovation Is Transforming How Audits Are Conducted-and Even What It Means to Be an Auditor. *Journal of Accountancy*, *223*(4), 28.

Ravuri, S., & Stolcke, A. (2016, March). A comparative study of recurrent neural network models for lexical domain classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on* (pp. 6075-6079). IEEE.

Russell, Stuart J.; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2*

Russell, Stuart J.; Norvig, Peter (2009). *Artificial Intelligence: A Modern Approach (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall. ISBN 0-13-604259-7.*

Rybach, D., Hahn, S., Lehnen, P., Nolden, D., Sundermeyer, M., Tüske, Z., ... & Ney, H. (2011, December). Rasr-the rwth aachen university open source speech recognition toolkit. In *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*.

Seide, F., Li, G., & Yu, D. (2011). Conversational speech transcription using context-dependent deep neural networks. In *Twelfth Annual Conference of the International Speech Communication Association*.

Seow, P. S. (2011). The effects of decision aid structural restrictiveness on decision-making outcomes. *International Journal of Accounting Information Systems*, *12*(1), 40-56.

Shivakumar, P. G., Li, H., Knight, K., & Georgiou, P. (2018). Learning from Past Mistakes: Improving Automatic Speech Recognition Output via Noisy-Clean Phrase Context Modeling. *arXiv preprint arXiv:1802.02607*.

Siau, K., & Rossi, M. (2011). Evaluation techniques for systems analysis and design modelling methods–a review and comparative analysis. *Information Systems Journal*, *21*(3), 249-268.

Stolcke, A., & Shriberg, E. (1996, May). Statistical language modeling for speech disfluencies. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on* (Vol. 1, pp. 405-408). IEEE.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2017). The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(2), 93.

Täckström, O., Das, D., Petrov, S., McDonald, R., & Nivre, J. (2013). Token and type constraints for cross-lingual part-of-speech tagging. *Transactions of the Association for Computational Linguistics*, *1*, 1-12.

Tang, R., & Lin, J. (2017). Honk: A PyTorch reimplementation of convolutional neural networks for keyword spotting. *arXiv preprint arXiv:1710.06554*.

Tur, G., & Hakkani-Tür, D. (2011). Human/human conversation understanding. *Spok. Lang. Underst.: Syst. for Extracting Semantic Inf. from Speech*, 225-255.

Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tur, D., Dowding, J., ... & Frederickson, C. (2010). The CALO meeting assistant system. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(6), 1601-1611.

US Census Bureau. (2017). North American industry classification system.

Venable, J., Pries-Heje, J., & Baskerville, R. (2012, May). A comprehensive framework for evaluation in design science research. In *International Conference on Design Science Research in Information Systems* (pp. 423-438). Springer, Berlin, Heidelberg.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European Journal of Information Systems*, *25*(1), 77-89.

Vera-Muñoz, S., Ho, J., & Chow, C. (2006). Enhancing knowledge sharing in public accounting firms. Accounting Horizons, 20(2), 133–155. doi:10.2308/acch.2006.20.2.133

Waibel, A., Stiefelhagen, R., Carlson, R., Casas, J., Kleindienst, J., Lamel, L., ... & Polymenakos, L. (2010). Computers in the human interaction loop. In *Handbook of Ambient Intelligence and Smart Environments* (pp. 1071-1116). Springer, Boston, MA.

WRDS Database. (2018). https://wrds-web.wharton.upenn.edu/wrds/

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., ... & Zweig, G. (2017, March). The Microsoft 2016 conversational speech recognition system. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on* (pp. 5255-5259). IEEE.

Yaman, S., Hakkani-Tür, D., & Tur, G. (2010). Social role discovery from spoken language using dynamic bayesian networks. In *Eleventh Annual Conference of the International Speech Communication Association*.

Zhang, S., Zheng, D., Hu, X., & Yang, M. (2015). Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation* (pp. 73-78).