2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Text Mining with Information Extraction for Chinese Financial Knowledge Graph

Yung-Wei Teng
Graduate Institute of Information Management
National Taipei University
New Taipei City, Taiwan
s711036115@gm.ntpu.edu.tw

Min-Yuh Day*
Graduate Institute of Information Management
National Taipei University
New Taipei City, Taiwan
myday@gm.ntpu.edu.tw

Pei-Tz Chiu
Graduate Institute of Information Management
National Taipei University
New Taipei City, Taiwan
s711036103@gm.ntpu.edu.tw

Abstract*—Financial Documents reveal important financial information about a company's financial performance which plays a vital role not only to the stakeholders but also to the public. Therefore, many researchers utilize dynamic Text mining methods in financial document to identify, analyze, predict or evaluate a company's future financial value. In order to find deeply the relationship between companies and the stakeholders, provide a simplified method for them to identify the future financial performance of the corporation. In this paper, we present a Chinese Information Extraction System (CFIES) for Financial Knowledge Graph (FinKG). The major findings of the research show an increased importance of the key audit matters in finance. The major research contribution of this paper is that we have developed CFIES which can extract the tuples from the financial reports. The adoption of the information system can assist the development of a knowledge graph that can discover deep financial knowledge in the finance domain. The managerial implication is that building CFIES can efficiently enable us to clarify the complicated relationship between the corporations, board of directors, investors, and especially the asset, assisting the stakeholders to discover a new financial knowledge representation and to make a financial decision.

***Keywords*—Text Mining, Information Extraction, Finance, Chinese Knowledge Graph, Key Audit Matters**

I. INTRODUCTION

In the regulated financial market, the corporation is obligated to disclose its official financial documents under the regulation of International Financial Reporting Standards (IFRS), namely financial report and financial statements. The financial statements, the balance sheet, Statement of Comprehensive Income, Statement of Cash Flow, Statements of Change in Equity and the Independents Auditor's (Review) Report are also mandatory included. Besides, the Independent Auditor's Report is composed of four parts: 1) Opinion, 2) Basis of Opinion, 3) Key Audit Matters (KAM) and 4) Responsibilities for Management and Those Charged with Governance for the Consolidate Financial Statements, which the Key Audit Matters points out the key factor of the financial influential factor and its financial risk exposure of the corporation. The research motivation is based on abovementioned sophisticated and time-consuming financial information. In addition, seldom of the research is illustrated about the importance of key audit matters in text-mining, specifically, in information extraction system.

In summary, due to the research motivation, it is essential to analyze the large-scale text from the financial documents, not to mention in Chinese. Therefore, the research objective is to establish Chinese Financial Information Extraction (CFIES) as a key component of financial knowledge graph (FinKG) and analyze the importance of key audit matters as the resource of financial knowledge graph construction.

The remaining section of the paper is organized as follows. The related work is introduced in Section 2. Section 3 describes the proposed FinKG architecture with the adoption of Chinese Financial Information Extraction System (CFIES). Data Analysis and Discussion of the key audit matters research is presented in Section 4. Finally, Section 5 provides the conclusion of the paper.

II. RELATED WORK

A. Information Extraction

Information Extraction(IE) is to convert the unstructured text into structured knowledge representation [8] and extract the relational tuples contained in the information. Sarhan and Spruit [9] compared three approaches of Open IE technique includes, machine-learning, hand-crafted rule based and neural network [9]; hence, in this study, we focus more on the neural network OIE approach.

B. Name Entity Recogonition

Entity and relation recognition is to identify the entities from the triples and categorized them based on different attributes e.g. person, location, organization, time etc. Liu et al. [10] reviewed six Chinese dataset of the name entity recognition task, and organized the name entity types.

There are several methods for NER, and the previous studies had shown that the neural-based NER model has been achieved the state-of-the-art results.

The previous studies had applied Long Shot Term Model (LSTM) with Conditional Random Fields (CRFs) [11]. However, with the raise popularity of encoder-decoder architecture and the attention mechanism [12], transformer-based model had been employed in many domain specific task. A prominent example of transformer-based model is BERT, is also the state-of-the-art language representation model. Xu et al. [6] exploit the Bio-NER for the construction of knowledge graph in biomedical domain based on BioBERT, in which can

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

recognize and discover the biomedical entities. Moreover, Yao et al. [13] proposed KG-BERT for knowledge graph completion by utilizing BERT, which had achieved the average accuracy 91.9% in both WordNet and Freebase dataset.

C. Relation Extraction

The main goal of Relation Extraction is to identify and link the relation between entities which extracted from the text. The methodologies can be classified as three main approaches Discourse-based, Distant Supervised-based, and OIE-based [14]. In this section, we specialize in the OIE-based approach.

The traditional Open IE-Based approach such as TEXTRUNNER [15] without predefined the relation, it may extract the different strings which represent the same relation; therefore, it can determine the synonymous relations and objects. However, this may cause the confusion the selection between subject and object, and the incoherent relation. In order to address the issue, the mapping relation phrase [16] comes out based on the relations constraint [17].

D. Knowledge Graph

Knowledge Graph is one of a downstream application of information extraction, which is also a novel method of knowledge representation, likewise it is an output of knowledge engineering in the field of Artificial Intelligent, which contains three important elements including entities, entity labels and relations [18].

With the emerging of knowledge graphs, there are several freely accessible Link Open Data with general knowledge and cross-domain knowledge graphs appear. They can be categorized in curation e.g., Freebase, and Wikidata; extraction from semi-structured or structured data e.g., YAGO and DBpedia; extraction from unstructured information e.g., NELL.

Freebase [19] is contributed by a great numbers of volunteers; hence, it has the largest number of triples includes entities (49M) and relations (70K). It has strong knowledge representation in media domain

Wikidata [20] a project of Wikipedia is also uses the manner of crowdsourcing. It stores the facts and the indicated sources of the facts so that it can be checked. Besides, it also imports a large scale of dataset such as the integration with Freebase.

DBpedia [21] is made of information from Wikipedia such as info-box and external links; therefore, all the pages in Wikipedia becomes the entities, on the other hands, the values from the pages present the attribute in the graph.

YAGO —Yet Another Great Ontology [22] is a multi-lingual knowledge graph consists of information extracted from Wikipedia, WordNet [23] and GeoNames. It provides the most amounts of classes and highest number of unique entities among the others with the small fraction of each class.

NELL— Never Ended Language Learning [24] works on a large scale of text form the Web via continuously-coupled process. It is originally trained on a few samples and continuously learn the text pattern to the indicated facts, likewise the same method to extract new entities and relations, so that it may extend its knowledge base.

TABLE 1 OVERVIEW OF KG APPROACH IN DOMAIN SPECIFIC

Ref.	Domain	Construction Algorithm(s)
[2]	Cybersecurity	NER: Regular Expression + CRF RE: Neural Network
[3]	Cybersecurity	OIE Model: BiGRU-Attention NER: BiGRU-CRF
[6]	Biomedical	BERT
[7]	Medical	NER: Bidirectional Maximum Matching (BMM)& BiLSTM-CRF

TABLE 2 OVERVIEW OF KNOWLEDGE GRAPH IN FINANCE

Ref	Domain	Construction Algorithm(s)	KG Resource(s)
[1]	Finance	OIE: OpenIE v5.13 RE: BiLSTM-Multi-head attention	Financial news from Chinese financial market
[4]	Finance	Annotated by Experts	Financial research reports
[5]	Finance	NER: BiLSTM-CRF	US Financial news dataset

E. Knowledge Graph in Domain Specific

Knowledge graphs have been widely applied in dynamic Natural Language Process [25] tasks such as question answering systems, recommender systems, information retrieval and specific domain like medical, cyber security, education and finance [26]. In this section, we focus on the knowledge graph construction in specific domain. Table 1 presents the overview of knowledge graph approach in domain specific.

With the raising demand of medical information from the public, the studies of reasoning medical knowledge have grown significantly. Xu et al. [6] build a PubMed Knowledge Graph by utilizing PubMed which is an important resource in the medical domain. However, it is ambiguous and hard to extract that raise the difficulty of knowledge discovery; therefore, in order to solve the issues, Xu et al. [6] developed Bio-BERT model [27] based on the Bidirectional Encoder Representations from Transformers (BERT) [28] for name entity extraction in biomedical domain to establish the PubMed KG. In addition, Yuan et al. [29] proposed a manner for the construction of biomedical knowledge graphs based on minimally supervised approach which is capable of reducing the weight of noisy instance. By adopting a piecewise CNN [30]with selective attention model in relation refinement, which is able to extract the open-ended relation with high precision, can reach an optimal performance and also be extended to other specific domain.

F. Knowledge Graph in Finance

Finance has been a promising field for knowledge graph application. The use of knowledge graph can help investors, stakeholders, employee and the publics to mine the deep knowledge in the sophisticated textual data. Table 2 presents the overview of knowledge graph in finance. In order to emphasize the importance of knowledge graph application in finance, Cheng et al. [1] present a financial knowledge graph by adopting OpenIE-based model for tuple extraction, and utilizing Bi-LSTM for relation extraction task which greatly improve the quantitative investment.

\

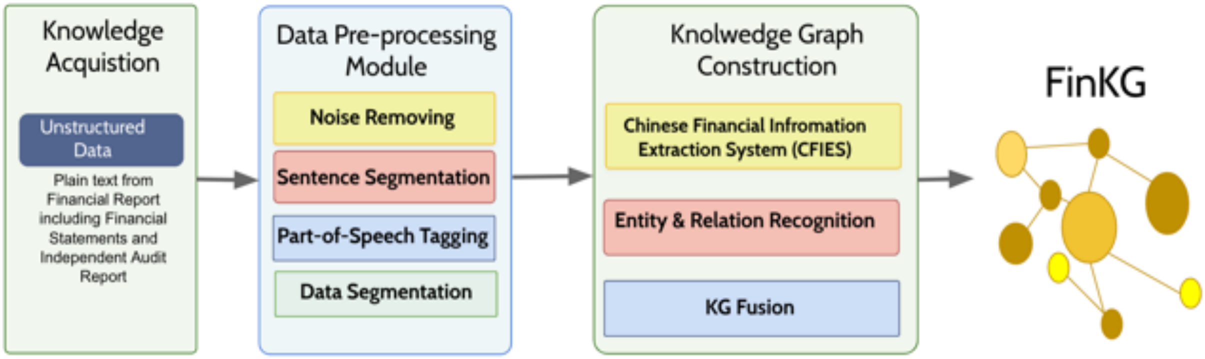

Fig. 1 The Proposed System Architecture of Financial Knowledge Graph (FinKG)

Furthermore, Wang et al. [4] conducted a financial research report knowledge graph (FR2KG) in Chinese, which is covered with 10 entity types and 19 relation types. The BERT-based model had been adopted for the Name Entity Recognition task in the construction of FR2KG. In respect of the relation extraction, the paper has reported that the co-occurrence approach had been utilized

G. Knowledge Graph Construction

Knowledge graphs is constructed in the form of triples with the head entity, relation and tail entity. Speaking of knowledge graphs construction, it is important to mention the relational forms of Resource Description Framework (RDF) by W3C which proposed a framework for information representation from the Web [31]. Knowledge Graph can be originally conducted under the concept of RDF — (subject, predicate, object) (SPO). Subject and object are corresponded to entities and the predicate represent the relation between them. After mapping multiple SPO triples together into the multigraph, where nodes indicate entities and the directional edge represents the relations, they become Knowledge Graphs (KG).

III. Research Methodology and System Architecture

A. Research Methodology

In this paper, the research methodology we adopt is the System Development Research Methodology [32]. According to Nunamaker Jr et al. [32], there are five main process of system development including (1)1 Construct a Conceptual Framework, (2) Develop a System Architecture, (3) Analyze and Design the System, (4) Build the System, (5) Observe and Evaluate the System.

1. *Construct a Conceptual Framework*
   Purpose the research question clearly, realize the process of building a Chinese Financial Information Extraction System (CFIES) with OpenIE-based model and review the relevant studies.
2. *Develop a System Architecture*
   Exploit the architecture of constructing the Chinese Financial Information Extraction System (CFIES) and define each stage and relationship between them.
3. *Analyze and Design the System*
   Design the database and knowledge base system for constructing financial knowledge graph.
4. *Build the System*
   Through building Chinese Financial Information Extraction System (CFIES) to realize the core concept, value, framework of the research and understand the complexity.
5. *Observe and Evaluate the System*
   Via the review of case studies and field research to observe the application of Chinese Financial Information Extraction System (CFIES) and evaluate through experimentation and observation to discover the new concept of the usage.

B. Proposed System Architecture

The proposed system architecture of financial knowledge graph (FinKG) is presented in Fig. 1. Our purposed Chinese Financial Information Extraction System (CFIES) is one of the steps of constructing the FinKG. First, we collect information from financial report, including Financial Statements and Independent Audit report. After that, we build a Data Pre-processing Module with sentence Segmentation, Post-of-Speech Tagging, Remove Noise and Data Segmentation. On completion of financial knowledge graph, the transformer-based model for Chinese open information extractor to output the triples. Finally, with the financial name entity and relation recognition model for identifying the entities and relations to construct the financial knowledge graph (FinKG) the overall architecture. The example of a set of triples in knowledge based is presented in Fig. 2. Fig. 3 presents the example of financial knowledge graph with entities and relations in knowledge graph.

Our research first mainly focuses on Key Audit Matters from Independent Audit Report. In order to construct a comprehensive financial knowledge graph (FinKG), key audit matter is one of the important information to realize the financial performance of the corporation. Therefore, we utilize the web-crawler to collect the Independent Audit Report of 50 corporation from FTSE TWSE Taiwan 50 Index (0050) from January, 1st, 2019 to December, 31st, 2021 in plain text in Financial Report from Market Observation Post System

<本會計師, Audit, 台積電 >
<The Accountant, Audit, TSMC>
<不動產、廠房及設備, Contain, 未完工程與待驗設備 >
< Property, Plant and Equipment, Audit, Unfinished construction and equipments pending acceptance >
<本會計師, Audit, 不動產、廠房及設備 >
<The Accountant, Audit, Property, Plant and Equipment >
<台積電, Disclose, 不動產、廠房及設備>
<TSMC, Disclose, Plant and Equipment >
<台積電, Disclose, 未完工程與待驗設備 >
<TSMC, Disclose, Unfinished construction and equipments pending acceptance >

Fig. 2 Example of a Set of Triples in Knowledge Base

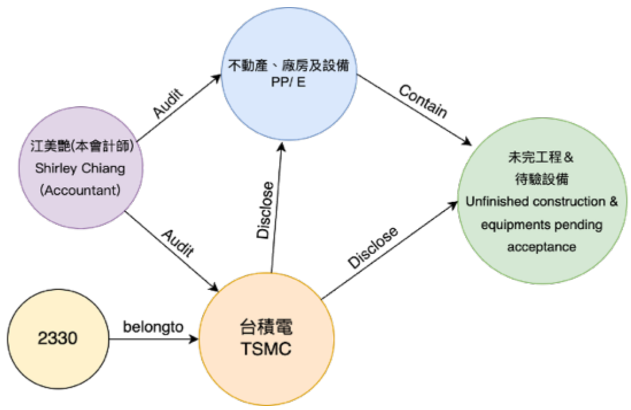

Fig. 3 Example of Financial Knowledge Graph with Entities and Relations in Knowledge Graph

(MOPS). After that, we use part-of-speech (POS) Tagging technique to select 30 relation types into the relation type pools. Finally, we adopt Bidirectional Encoder Representations from Transformers (BERT) to develop the CFIES, turn unstructured textual data to the triples.

IV. Data Analysis and Discussion

In this section, we conduct the trend analysis, keywords analysis, region analysis, and published journals analysis on Key Audit Matter (KAM) research from Scopus for constructing the architecture of Chinese Financial Information Extraction System (CFIES).

A. *Trend Analysis*

Fig. 4 reveals the trend of the research about Key Audit Matters on Scopus from 1978 to 2022. Although the number of research is likely to remain steady before 2000, the graph shows there has been a slightly increased until 2017, and grows sharply after 2017 until now. The first study about Key Audit Matters appeared in 1978. With the importance of audit information raise, more than 66% research has been published after 2017. The research on Key Audit Matters peaked in 2022, which takes approximately 17% of overall and expected to keep increasing in the future.

B. *Keyword Analysis*

Table 3 shows the keyword analysis for the key audit information in Natural Language Process (NLP) domain. As shown in the table, there are 242 studies on Key Audit Matters. And around 60 research is about Key Audit Matters disclosure, however only 1 study contains text mining and Key Audit matter. Therefore, we can see that text mining in key audit matters is worth of the researchers, accountants and investors for further research. TABLE shows the frequency and average citation of co-occurrence keyword in Key Audit Matters research from Scopus database. There are roughly 672 keywords in 242 research for Key Audit Matters. The top 20 keywords are Key audit matter(s), Audit Report, Auditing, Audit Quality, Audit, Corporate Governance, Thailand, Critical Audit Matters, ISA701, Readability, Quality, External Audit, COVID-19, South Africa, UK, Regulation, Internal Auditing, Audit Quality, Auditor Liability and Audit Committee(s).

Fig. 5 visualizes the co-occurrence keyword with around 12 cluster in KAM research. We can see the cluster in color of pink, it suggests that risk disclosure and textual analysis are highly

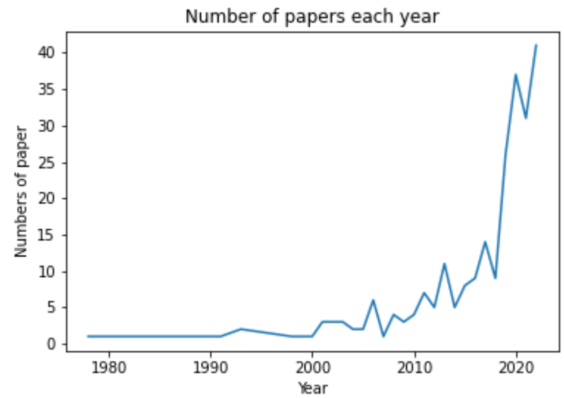

Fig. 4 Key Audit Matter (KAM) Trend Analysis

Table 3 Keyword Statistic

Keyword	Number of Studies
Key Audit Matters	242
Key Audit Matters, Disclosure	60
Financial Audit, Information Extraction	30
Audit Information, Knowledge Graph	30
Key Audit Matter, Text Mining	1

Table 4 Frequency of co-occurrence keyword in KAM research

Rank	Keyword	Frequency	Percentage	Citation (avg)
1	Key audit matter(s)	76	11.3%	17
2	Audit report	26	3.9%	40
3	Auditing	11	1.6%	3.7
4	Audit quality	11	1.6%	7
5	Audit	11	1.6%	2.07
6	Corporate governance	10	1.5%	16.9
7	Thailand	6	0.9%	3.67
8	Critical audit matters	5	0.7%	20.42
9	ISA 701	4	0.6%	14
10	Readability	4	0.6%	5.2
11	Quality	4	0.6%	17.75
12	External audit	4	0.6%	8.5
13	COVID-19	4	0.6%	2.67
14	South Africa	4	0.6%	17
15	UK	3	0.4%	8.67
16	Regulation	3	0.4%	4.33
17	Internal auditing	3	0.4%	17
18	Audit Quality	3	0.4%	7.05
19	Auditor liability	3	0.4%	9.5
20	Audit committee(s)	3	0.4%	74.3

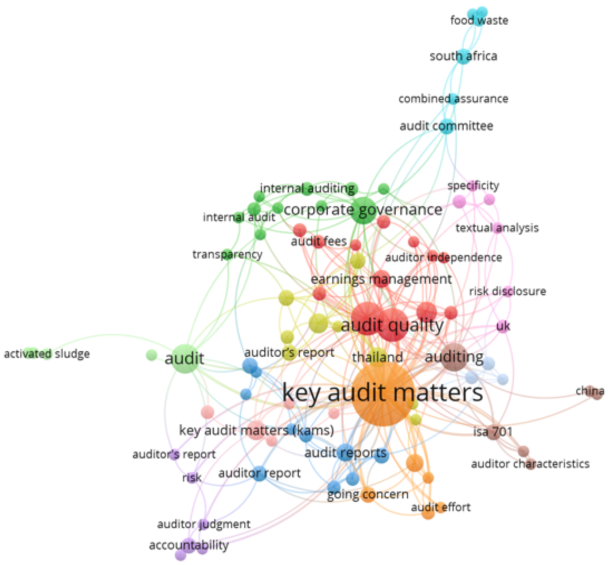

Fig. 5 Co-occurrence Keyword Visualization for KAM Research

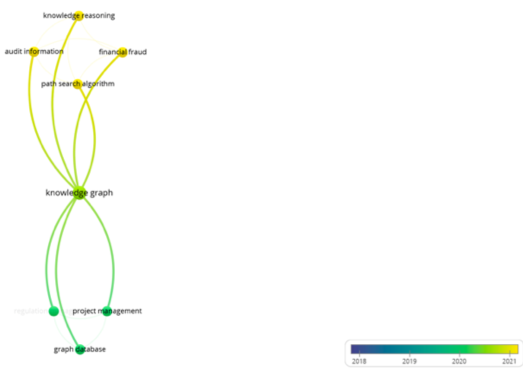

Fig. 6 Co-occurrence Keyword Visualization for Audit Information Research

co-occurrence with Key Audit Matters, which is one of the most important keywords in Key Audit Matters and AI research.

Fig. 6 shows the visualization of the co-occurrence keyword in audit information and knowledge graph research which maps in Table 3. It reveals highly co-occurrence keyword of financial fraud and knowledge reasoning in audit information and knowledge graph research. Therefore, it highlights the importance of knowledge inference in audit information.

C. *Region Analysis*

Table 5 presents the top 20 country from the research affiliation in KAM research. The top 3 country in KAM research comes from United States, United Kingdom and Australia. However, Taiwan earned the twenty-fifth place among the 45 countries. The region analysis shows that Thailand, South Africa, and the United Kingdom are crucial countries for audit research, and there is a clear trend of increased research in these countries.

D. *Published Journal Analysis*

Table 6 lists the top 20 journals published Key Audit Matter research on Scopus. There are International Journal of Auditing, Managerial Auditing Journal, Water Science and Technology, Auditing, European Accounting Review, Journal of Applied Accounting Research, Institution of Chemical Engineers Symposium Series, British Accounting Review, Accounting Horizons, Accounting in Europe, Meditari Accountancy Research, International Journal of Disclosure and Governance, Pacific Accounting Review, Revista Contabilidade e Financas, Journal of Public Budgeting, Accounting and Financial Management, Revista de Contabilidad-Spanish Accounting Review, Cogent Business and Management, Business Horizons, Corporate Ownership and Control, PLoS ONE.

TABLE 5 TOP 20 AFFILIATION COUNTRY IN KAM RESEARCH

Rank	Country	Count	Percentage
1	United States	39	17%
2	United Kingdom	26	11%
3	Australia	26	11%
4	Thailand	10	4%
5	Germany	9	4%
6	South Africa	9	4%
7	Netherlands	8	3%
8	Spain	8	3%
9	China	8	3%
10	Malaysia	7	3%
11	Portugal	6	3%
12	Italy	5	2%
13	Canada	5	2%
14	Finland	4	2%
15	Poland	3	1%
16	Egypt	3	1%
17	Indonesia	3	1%
18	Denmark	3	1%
19	Romania	3	1%
20	New Zealand	3	1%

TABLE 6 TOP 20 JOURNAL PUBLISHED KEY AUDIT MATTER RESEARCH ON SCOPUS

Rank	Journal	Count	Percentage
1	International Journal of Auditing	12	4.96%
2	Managerial Auditing Journal	11	4.55%
3	Water Science and Technology	4	1.65%
4	Auditing	4	1.65%
5	European Accounting Review	4	1.65%
6	Journal of Applied Accounting Research	3	1.24%
7	Institution of Chemical Engineers Symposium Series	3	1.24%
8	British Accounting Review	3	1.24%
9	Accounting Horizons	3	1.24%
10	Accounting in Europe	3	1.24%
11	Meditari Accountancy Research	3	1.24%
12	International Journal of Disclosure and Governance	3	1.24%
13	Pacific Accounting Review	3	1.24%
14	Revista Contabilidade e Financas	3	1.24%
15	Journal of Public Budgeting, Accounting and Financial Management	2	0.83%
16	Revista de Contabilidad-Spanish Accounting Review	2	0.83%
17	Cogent Business and Management	2	0.83%
18	Business Horizons	2	0.83%
19	Corporate Ownership and Control	2	0.83%
20	PLoS ONE	2	0.83%

V. Conclusion

Information extraction has been generated in different domain such as biomedical, cybersecurity and finance. Therefore, the major finding of the research is that research related to key audit matters has been increased its importance especially in United States and United Kingdom which correspond to our research objective. Additionally, after systematically review the research related to key audit matters, we discover that key audit matter usually accompanied with the audit quality and its liability.

The major research contribution of this paper is that we have developed Chinese Financial Information Extraction System (CFIES) as an important module for constructing the FinKG. The purposed CFIES may assist the contribution of the financial knowledge graph to infer the knowledge between entities and relations. The managerial implication of the research is for investors and management of the corporation to discover a new financial knowledge representation between entities and to make the more efficient financial decision.

Further research into the financial statements and their notes can be added to part of the Chinese Financial Knowledge Graph, which may develop as one of the biggest financial knowledge graphs in the Chinese dataset.

Acknowledgment

This research was supported in part by Ministry of Science and Technology (MOST) under the grant number 110-2410-H305-013-MY2 and National Taipei University (NTPU) under the grant number 111-NTPU-ORDA-F-001 and 111-NTPUORDA-F-003.

References

[1] D. Cheng, F. Yang, X. Wang, Y. Zhang, and L. Zhang, "Knowledge graph-based event embedding framework for financial quantitative investments," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 2221-2230.

[2] A. Piplai, S. Mittal, A. Joshi, T. Finin, J. Holt, and R. Zak, "Creating cybersecurity knowledge graphs from malware after action reports," *IEEE Access,* vol. 8, pp. 211691-211703, 2020.

[3] I. Sarhan and M. Spruit, "Open-CyKG: An Open Cyber Threat Intelligence Knowledge Graph," *Knowledge-Based Systems,* p. 107524, 2021.

[4] W. Wang, Y. Xu, C. Du, Y. Chen, Y. Wang, and H. Wen, "Data set and evaluation of automated construction of financial knowledge graph," *Data Intelligence,* vol. 3, no. 3, pp. 418-443, 2021.

[5] S. Elhammadi, "Financial knowledge graph construction," University of British Columbia, 2020.

[6] J. Xu *et al.*, "Building a PubMed knowledge graph," *Scientific data,* vol. 7, no. 1, pp. 1-15, 2020.

[7] L. Li *et al.*, "Real-world data medical knowledge graph: construction and applications," *Artificial intelligence in medicine,* vol. 103, p. 101817, 2020.

[8] D. Jurafsky, *Speech & language processing*. Pearson Education India, 2000.

[9] I. Sarhan and M. Spruit, "Uncovering algorithmic approaches in open information extraction: a literature review," in *30th Benelux Conference on Artificial Intelligence*, 2018: Springer CSAI/JADS, pp. 223–234.

[10] P. Liu, Y. Guo, F. Wang, and G. Li, "Chinese named entity recognition: The state of the art," *Neurocomputing,* vol. 473, pp. 37-53, 2022.

[11] G. Kim, C. Lee, J. Jo, and H. Lim, "Automatic extraction of named entities of cyber threats using a deep Bi-LSTM-CRF network," *International journal of machine learning cybernetics,* vol. 11, no. 10, pp. 2341-2355, 2020.

[12] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998-6008.

[13] L. Yao, C. Mao, and Y. Luo, "KG-BERT: BERT for knowledge graph completion," *arXiv preprint arXiv:.03193,* 2019.

[14] J. L. Martinez-Rodriguez, I. Lopez-Arevalo, and A. B. Rios-Alvarado, "Openie-based approach for knowledge graph construction from text," *Expert Systems with Applications,* vol. 113, pp. 339-355, 2018.

[15] A. Yates, M. Banko, M. Broadhead, M. J. Cafarella, O. Etzioni, and S. Soderland, "Textrunner: open information extraction on the web," in *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 2007, pp. 25-26.

[16] A. Dutta, C. Meilicke, and H. Stuckenschmidt, "Enriching structured knowledge with open information," presented at the Proceedings of the 24th international conference on world wide web, 2015.

[17] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011, pp. 1535-1545.

[18] J. Pujara, H. Miao, L. Getoor, and W. Cohen, "Knowledge graph identification," in *International semantic web conference*, 2013: Springer, pp. 542-557.

[19] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 1247-1250.

[20] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM,* vol. 57, no. 10, pp. 78-85, 2014.

[21] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "Dbpedia: A nucleus for a web of open data," in *The semantic web*: Springer, 2007, pp. 722-735.

[22] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A large ontology from wikipedia and wordnet," *Journal of Web Semantics,* vol. 6, no. 3, pp. 203-217, 2008.

[23] G. A. Miller, *WordNet: An electronic lexical database*. MIT press, 1998.

[24] A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka Jr, and T. M. Mitchell, "Coupled semi-supervised learning for information extraction," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 101-110.

[25] C. Jia, Y. Shi, Q. Yang, and Y. Zhang, "Entity enhanced BERT pre-training for Chinese NER," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 6384-6396.

[26] X. Zou, "A survey on application of knowledge graph," in *Journal of Physics: Conference Series*, 2020, vol. 1487, no. 1: IOP Publishing, p. 012016.

[27] M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, and U. Leser, "Deep learning with word embeddings improves biomedical named entity recognition," *Bioinformatics,* vol. 33, no. 14, pp. i37-i48, 2017.

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:.04805,* 2018.

[29] J. Yuan *et al.*, "Constructing biomedical domain-specific knowledge graph with minimum supervision," *Knowledge Information Systems,* vol. 62, no. 1, pp. 317-336, 2020.

[30] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural relation extraction with selective attention over instances," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 2124-2133.

[31] G. Klyne, "Resource description framework (RDF): Concepts and abstract syntax," http://www.w3.org/TR//REC-rdf-concepts-20040210/, 2004.

[32] J. F. Nunamaker Jr, M. Chen, and T. D. Purdin, "Systems development in information systems research," *Journal of management information systems,* vol. 7, no. 3, pp. 89-106, 1990.