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 for Financial reports 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 an information extraction system for financial knowledge graph. We develop an AI natural language processing model and extract the tuples from the financial report in the composition of an unstructured large corpus. As the result, building a Chinese Financial Information Extraction System (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. The adoption of the information system can assist the development of a knowledge graph that can discover deep financial knowledge in the Finance Domain.

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

I. Introduction

In the past decades, the rapidly growth of advanced technology has led to the age of information explosion, in other words, the desire for knowledge has being increased. Therefore, the studies of knowledge management springs up to effectively address the instant information. At the same time, the need of instant information, such as numbers, characters, words and text, becomes vital for the public, enterprise, investors and even government.

According to Tanwar, Duggal and Khatri [1], there are over 95% of information belong to unstructured data; thus, the research related to processing the unstructured data becomes a key instrument for the society. Nevertheless, manually handle a large scale of unstructured information is sophisticated and time-consuming, not to mention that we might loss valuable information within the text, thus text mining is one of an important task of Natural Language Process for addressing this issue. Under the circumstance, fast information extraction is one of the popular issues in society.

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 information disclosed in financial report contains the prosecution strategy of the corporation, Summary of Significant Accounting Policies, Financial Risk Management, and list of financial instruments etc. As for 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 mandatory included. The abovementioned financial information is complex, not easy to understand and even time-consuming to read.

In summary, with the aforementioned description, it is essential to analyze the large-scale text from the disclosed documents, not to mention in Chinese. Therefor, financial information extraction is a key method for text mining in finance.

The remaining section of the paper is organized as follows. The related work is introduced in Section 2. Section 3 describes the proposed system architecture. Data Analysis and Discussion of the key audit matters research is presented in Section 4. Finally, Section 5 is the conclusion and discussion.

II. Related Work

A. Information Extraction

Information Extraction(IE) is to convert the unstructured text into structured knowledge representation [2] and extract the relational tuples contained in the information. Sarhan and Spruit [3] compared three approaches of Open IE technique includes, machine-learning, hand-crafted rule based and neural network [3]; 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, Guo, Wang and Li [4] 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.

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

The previous studies had applied Long Shot Term Model (LSTM) with Conditional Random Fields (CRFs)[5]. However, with the raise popularity of encoder-decoder architecture and the attention mechanism[6], 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, Kim, Song, Jeong, Kim, Kang, Rousseau, Li, Xu and Torvik [7] exploit the Bio-NER for the construction of knowledge graph in biomedical domain based on BioBERT, in which can recognize and discover the biomedical entities. Moreover, Yao, Mao and Luo [8] 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.[9]. In this section, we specialize in the OIE-based approach.

The traditional Open IE-Based approach such as TEXTRUNNER [10] 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[11] comes out based on the relations constraint [12].

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 [13].

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 [14] 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 [15] 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 [16] is made of information from Wikipedia such as infobox 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 [17] is a multi-lingual knowledge graph consists of information extracted from Wikipedia, WordNet[18] and GeoNames. It provides the most amounts of classes and highest number of unique entities among the others with the small fraction of each classes.

NELL— Never Ended Language Learning [19] 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.

E. Knowledge Graph in Domain Specific

Knowledge graphs have been widely applied in dynamic Natural Language Process [20] tasks such as question answering systems, recommender systems, information retrieval and specific domain like medical, cyber security, education and finance [21]. In this section, we focus on the knowledge graph construction in specific domain. (Table 2)

TABLE 1 OVERVIEW OF KG APPROACH IN DOMAIN SPECIFIC

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

Source : Adapted from Abu-Salih [25]

With the raising demand of medical information from the public, the studies of reasoning medical knowledge have grown significantly. Xu, Kim, Song, Jeong, Kim, Kang, Rousseau, Li, Xu and Torvik [7] 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, Kim, Song, Jeong, Kim, Kang, Rousseau, Li, Xu and Torvik [7] developed Bio-BERT model[26] based on the Bidirectional Encoder Representations from Transformers (BERT)[27] for name entity extraction in biomedical domain to establish the PubMed KG. In addition, Yuan, Jin, Guo, Jin, Zhang, Smith and Luo [28] 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 [29]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. In order to emphasize the importance of knowledge graph application in

\

finance, Cheng, Yang, Wang, Zhang and Zhang [30] 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.(Table 3)

TABLE 2 OVERVIEW OF KNOWLEDGE GRAPH IN FINANCE

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

Source: This Study

Furthermore, Wang, Xu, Du, Chen, Wang and Wen [31] conducted a financial research report knowledge graph (FR2KG) in Chinese, which is covered with 10 entity types (Table 4) and 19 relation types (Table 5). 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 Construcution*

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 [33]. 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). The example of knowledge graph is illustrated in figure 1.

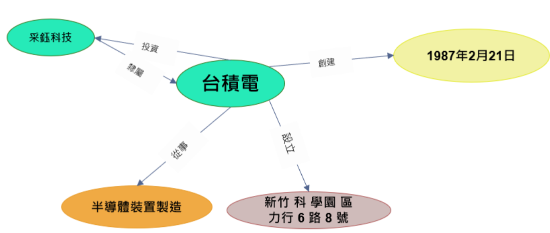

Fig. 1.Example of Financial Knowledge Graph with entities and relations in knowledge graph

(台積電,設立, 1987年2月21日)
(台積電,設立,新竹 科 學園 區力行6路8號)
(台積電,從事,半導體裝置製造)
(台積電, 投資, 采鈺科技)
(采鈺科技,隸屬, 台積電)

Fig. 2.Example of a set of triples in knowledge base

III. RESEARCH METHODOLOGY AND SYSTEM ARCHITECTURE

A. *Research Methodology*

In this paper, the research methodology we adopt is the System Development Research Methodology [34]. According to Nunamaker Jr, Chen and Purdin [34], 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 stages 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 Research Architecture*

Our purposed Chinese Financial Information Extraction System (CFIES) is one of the steps of constructing the financial knowledge graph, which is shown in the below framework (Figure 3.). First, we collect the data via the financial report, building 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.

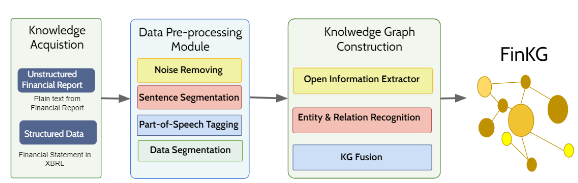

Fig. 3.The research architecture of the study

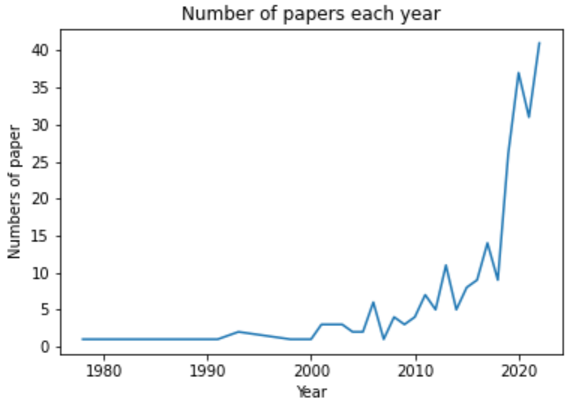

Fig. 4.Key Audit Matter (KAM) Trend Analysis

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 for constructing the architecture of Chinese Financial Information Extraction System (CFIES) from Scopus.

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 is the keyword statistic for the key audit information in Natural Language Process (NLP) domain. As shown in the table, there are around 242 studies are about Key Audit Matters. And around 60 research is about Key Audit Matters disclosure, however only 1 study contains text mining. Therefore, we can see that text mining in key audit matters is worth of the researchers, accountants and investors for further research.

Table 4 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 co-occurrence with Key Audit Matters, which is one of the most important keyword in Key Audit Matters and AI research.

Table 3. Keyword Statistic

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

Table 4 Frequency of co-occurrence keyword in KAM research

Rank	Keyword	Frequency	Percentage	Average Citation
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%	2.6667
14	South Africa	4	0.6%	17
15	UK	3	0.4%	8.6667
16	Regulation	3	0.4%	4.3333
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

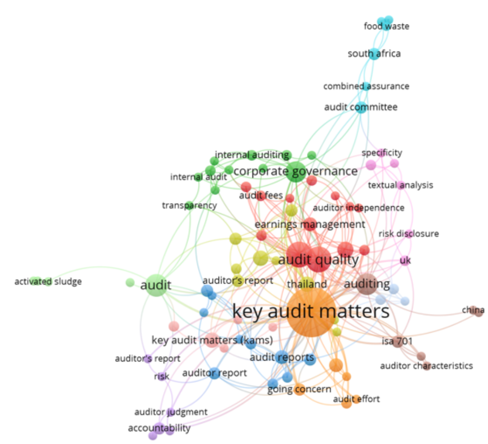

Fig. 5. Co-occurrence Keyword Visualization for KAM Research

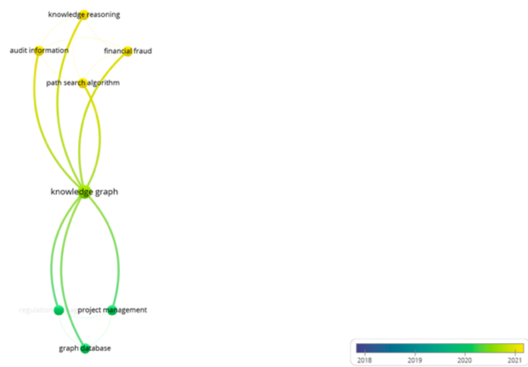

Fig. 6.Co-occurrence Keyword Visualization for Audit Information Research

TABLE 5.COUNT OF 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%
21	Brazil	3	1%
22	Switzerland	3	1%
23	Norway	3	1%
24	Taiwan	2	1%
25	South Korea	2	1%

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.

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%

C. Region Analysis

Table 5 the statistic of the country from research affiliation. The top 3 comes from United States, United Kingdom and Australia. However, Taiwan earned the twenty-fifth place among the 45 countries. From the data in Table 5 and Table 4, it is apparent that the Thailand, South Africa and United Kingdom are essential countries for the research in key audit matters and also is the clear trend of increasing.

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.

V. CONCLUSION

Information extraction has been generated in different domain such as biomedical, cybersecurity and financial. In this paper, we introduce Chinese Information Extraction System (CIFS), which is an important module for constructing the FinKG, which can be used for developing the financial question and answering system with adding the query system, or being applied for financial fraud detection task.

Furthermore, the developed triples set will be published in Github, an open resource platform, for the public to create and edit more Chinese financial reports of other Taiwanese company. So that the dataset of tuples will not just be limited in the corporation of FTSE TWSE Taiwan 50 Index (0050), but become an open data for other researchers to study in the domain of finance. In addition, it will be presented as one of the Chinese Financial Report datasets and established for the further research.

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] M. Tanwar, R. Duggal, and S. K. Khatri, "Unravelling unstructured data: A wealth of information in big data." pp. 1-6.

[2] D. Jurafsky, Speech & language processing: Pearson Education India, 2000.

[3] I. Sarhan, and M. Spruit, "Uncovering algorithmic approaches in open information extraction: a literature review." pp. 223–234.

[4] 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.

[5] 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 learningcybernetics, vol. 11, no. 10, pp. 2341-2355, 2020.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need." pp. 5998-6008.

[7] J. Xu, S. Kim, M. Song, M. Jeong, D. Kim, J. Kang, J. F. Rousseau, X. Li, W. Xu, and V. I. Torvik, “Building a PubMed knowledge graph,” Scientific data, vol. 7, no. 1, pp. 1-15, 2020.

[8] L. Yao, C. Mao, and Y. Luo, “KG-BERT: BERT for knowledge graph completion,” arXiv preprint arXiv:.03193, 2019.

[9] 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.

[10] A. Yates, M. Banko, M. Broadhead, M. J. Cafarella, O. Etzioni, and S. Soderland, "Textrunner: open information extraction on the web." pp. 25-26.

[11] A. Dutta, C. Meilicke, and H. Stuckenschmidt, “Enriching structured knowledge with open information,” in Proceedings of the 24th international conference on world wide web, 2015, pp. 267-277.

[12] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction." pp. 1535-1545.

[13] J. Pujara, H. Miao, L. Getoor, and W. Cohen, "Knowledge graph identification." pp. 542-557.

[14] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge." pp. 1247-1250.

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

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

[17] 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.

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

[19] A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka Jr, and T. M. Mitchell, "Coupled semi-supervised learning for information extraction." pp. 101-110.

[20] C. Jia, Y. Shi, Q. Yang, and Y. Zhang, "Entity enhanced BERT pre-training for Chinese NER." pp. 6384-6396.

[21] X. Zou, "A survey on application of knowledge graph." p. 012016.

[22] 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.

[23] I. Sarhan, and M. Spruit, “Open-CyKG: An Open Cyber Threat Intelligence Knowledge Graph,” Knowledge-Based Systems, pp. 107524, 2021.

[24] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T.-H. Chang, and S. Wang, “Real-world data medical knowledge graph: construction and applications,” Artificial intelligence in medicine, vol. 103, pp. 101817, 2020.

[25] B. Abu-Salih, “Domain-specific knowledge graphs: A survey,” Journal of Network Computer Applications, vol. 185, pp. 103076, 2021.

[26] 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.

[27] 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.

[28] J. Yuan, Z. Jin, H. Guo, H. Jin, X. Zhang, T. Smith, and J. Luo, “Constructing biomedical domain-specific knowledge graph with minimum supervision,” Knowledge Information Systems, vol. 62, no. 1, pp. 317-336, 2020.

[29] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural relation extraction with selective attention over instances." pp. 2124-2133.

[30] D. Cheng, F. Yang, X. Wang, Y. Zhang, and L. Zhang, "Knowledge graph-based event embedding framework for financial quantitative investments." pp. 2221-2230.

[31] 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.

[32] S. Elhammadi, “Financial knowledge graph construction,” University of British Columbia, 2020.

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

[34] 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.