

Retraction

Retracted: Information Extraction and Knowledge Graph Construction for Enterprises in China's Free Trade Zone

Security and Communication Networks

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] L. Meng, "Information Extraction and Knowledge Graph Construction for Enterprises in China's Free Trade Zone," *Security and Communication Networks*, vol. 2022, Article ID 2962545, 7 pages, 2022.

Research Article

Information Extraction and Knowledge Graph Construction for Enterprises in China's Free Trade Zone

LingYan Meng ^{1,2}

¹Universiti Pendidikan Sultan Idris (UPSI), Faculty of Management and Economics, Tanjung Malim 35900, Perak Darul Ridzuan, Malaysia

²Yangdong(Beijing)Education & Technology CO, LTD, Beijing 100083, China

Correspondence should be addressed to LingYan Meng; menglingyan@yangdong-edu.com

Received 1 April 2022; Accepted 20 May 2022; Published 10 June 2022

Academic Editor: Muhammad Arif

Copyright © 2022 LingYan Meng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the establishment and rapid development of the China (Shanghai) Pilot Free Trade Zone (FTZ), the scale of enterprises in the zone has grown rapidly. This paper takes the actual needs of the Shanghai FTZ as the background, extracts key information such as entities from the Big Data related to Internet enterprises, further constructs the enterprise knowledge graph, and applies it to the supervision and service of the FTZ. The enterprise knowledge graph is constructed using the Neo4j graph database. To verify the effectiveness of the named entity identification and relationship extraction methods proposed in this paper, experiments were conducted to validate them, and both achieved good results.

1. Introduction

China (Shanghai) Pilot Free Trade Zone (“Shanghai FTZ”) is China’s pilot project for economic globalisation. The Shanghai FTZ adopts special regulatory policies and preferential taxation, including actively promoting cross-border RMB business, establishing a more liberal taxation policy, creating an open financial operating environment, and ensuring the efficiency of administration. The FTAZ has been a great catalyst for domestic re-export and offshore trade [1]. However, the information in the platform is not rich enough to provide comprehensive supervision and services to enterprises, so the FTZ urgently needs Internet Big Data and related technologies to support and strengthen the comprehensive supervision and services to enterprises [2].

The Chinese Academy of Communication’s (CAICT) China Big Data Development Survey Report shows that the size of China’s Big Data industry was RMB 470 billion in 2017, with an annual growth rate of over 30% [3]. Enterprises are a key to the development of Big Data as both the promoters and beneficiaries of Big Data. According to CAICT, nearly 2/3 of enterprises have already set up relevant

data analysis departments, and nearly 40% of enterprises have already used Big Data [4]. The application of Big Data has helped enterprises to successfully achieve intelligent decision-making, improve their operational efficiency, and reduce related management risks, among others.

Therefore, in the context of a commissioned project and the actual needs of the Shanghai Free Trade Zone, this paper will carry out Information Extraction (IE) from enterprise Big Data [5], which is a way to structure heterogeneous and unstructured data to produce organised information. In this paper, information extraction is carried out by two aspects: named entity recognition (NER) [6] and relation extraction (RE) [7] for enterprise Big Data. Named entity recognition is an integral part of natural language processing techniques such as event detection and machine translation [8]. For example, in event detection, such information as personal, place, and time is the necessary elements that constitute an event. This paper therefore first investigates how to extract key information such as personal, place, business name, and date from enterprise Big Data.

Relation extraction is likewise another important subtask of information extraction in natural language processing, which aims to extract semantic relations between entities

from text. Relation extraction also has an important role in question answering, drug informatics, and ontology learning [9]. On the basis of named entity recognition, this paper further investigates how to extract business-to-business and individual-to-business relationships as a way to better grasp information about individuals as well as business associations [10].

In order to provide a comprehensive and intuitive presentation of dynamic information about enterprises, the paper takes advantage of the structured knowledge of the Knowledge Graph (KG) [11] and finally uses the extracted entities and relationships to build a knowledge graph of enterprises, providing a comprehensive and intuitive presentation of dynamic information about enterprises in the FTZ. The Knowledge Graph was released by Google on 16 May 2012 and the essence of the Knowledge Graph is “things no strings” [12].

In summary, this paper takes the actual need of Shanghai FTZ as the background, identifies named entities and extracts relationships from company announcements, further constructs the enterprise knowledge graph, and finally deploys the enterprise knowledge graph application into the platform project. The implementation of the enterprise knowledge graph provides the management of the FTZ with comprehensive and intuitive information on enterprise dynamics, which can further establish the foundation for the upper-level services based on the enterprise knowledge graph.

2. Related Work

When faced with large amounts of data, the topic model obtains a textual representation by modelling the implied topics of the text, which is a significant improvement over traditional TF-IDF representation methods [13]. In the study of topic models, Latent Dirichlet Allocation (LDA) topic models are popular because of their simple parameters and thus less likely to be overheated. Therefore, in the data acquisition stage, this paper adopts an improved LDA-based topic model for initial screening of large amounts of text data [14].

The authors in [8] used a Gaussian function feature word weighting method to reduce the impact of high-frequency words on the topic distribution. Thus, solving the problem that deactivated words cannot be directly removed when the topic model is applied to a specific domain. The authors in [15] proposed a weakly supervised classifier, Classify-LDA [16], in which a professional in the relevant field assigns one or more category labels to each topic based on the words, it is most likely to contain during data annotation, and then it uses LDA to generate the topic model. A new LDA framework [17], called semantic similarity LDA (SS-LDA), was proposed in [18], which integrates common sense-based semantic similarity computation into the word distribution computation in LDA models and achieves a shift from syntax to semantics in topic analysis by gradually changing hyperparameters during computation.

Improvements to the LDA subject model have focused on the computation of two distributions, making the model

more task-appropriate for modelling text by overcoming the negative effects of statistical frequency.

Conditional Random Fields (CRFs) were proposed in [19] for application to annotation problems [20], where they proposed conditional random fields by changing the local normalisation used in maximum entropy Markov models to global normalisation in the computation of probability distribution. Subsequently, conditional random fields began to be widely studied in various fields. The authors in [21] used conditional random fields with a variety of traditional and novel features to identify classes of entities such as proteins and cell types, and the authors in [22] used conditional random fields to detect the boundaries of named entities.

The architecture of [23] using a hybrid of BiLSTM and CNN builds models based on BiLSTM and CRF. Unlike English, the composition of China is more complex, with not only a semantic distinction between characters and words, but also additional information such as strokes and radicals, which can significantly improve the semantic representation of the model.

In addition, the most common research approach is to integrate character embedding with word embedding, such as the hybrid word embedding approach proposed by [24] and the proposed integrating embedding method CWPC_BiAtt that introduce an attention mechanism. For a more spoken corpus such as the Weibo dataset, the authors in [25] proposed to ME-CNER to derive the character embedding of named entities in Chinese text.

3. Overview of the Methodology Framework

Figure 1 shows the architecture of the method we use, as follows.

As the enterprise domain is oriented for naming entities, the FTA has provided a partial list of enterprises, so in order to make effective use of the existing knowledge, the lexicon-based approach is first used to identify enterprise names from listed company announcements using a forward maximum matching algorithm. Although the dictionary-based approach can effectively identify some of the corporate names in the announcements, the dictionary approach cannot fully identify all the entities needed, so a deep learning-based approach is further proposed to identify names, places, companies, and dates, and then the results of the joint identification of the two methods are compared with entity annotation to obtain the final named entity identification results. The lexicon-based method ensures the correct recognition rate of existing business names to a certain extent, while the deep learning-based method can recognize more unknown entities, so this joint method has better results than the single method for named entity recognition [26].

Each module in the architecture process will be briefly outlined below.

The data in the data input section are obtained from the Internet using crawlers. The data in the announcements of

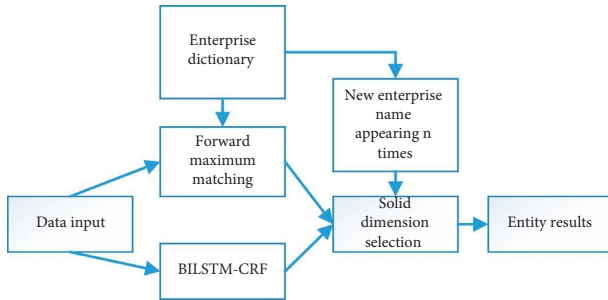


FIGURE 1: Architecture of the joint lexicon- and deep learning-based named entity recognition approach.

listed companies are unstructured data, which need to be first divided into sentences before being used as the original corpus for data input.

The lexicon-based recognition module is implemented by the lexicon-based forward maximal matching algorithm, which continuously intercepts characters of a certain length and matches them with existing company names in the lexicon, and when a company name in the lexicon is matched, the segment will be annotated.

This part is different from the lexicon method which can directly identify entities. The deep learning method first needs to train the model. If the BiLSTM-CRF model is not well trained, the model needs to be trained first. In the pre-processing part, the article needs to be divided into sentences first, and the annotation of the person, place, enterprise name, and date of the entity is carried out by the manual annotation method. After the model was trained, the listed company announcements were directly cut into sentences and input to the BiLSTM-CRF model named entity recognition.

After the named entity recognition based on the lexicon and the deep learning model, it is necessary to compare the company names identified by this joint approach to select the appropriate annotation result, as two annotation results are output correspondingly. In this section, we discuss the annotation selection of named entities for different scenarios. In addition, we also align the existence of business name abbreviations and full names in the annotation selection, while adding new business names to the lexicon that meets the recognition count threshold.

Once the entity annotation selection is complete, the annotation of the person, place, business name, and date in the sentence of the listed company announcement is obtained, and the corresponding entity is output through the annotation.

4. Enterprise Knowledge Graph Construction and System Application Implementation

Our work performs enterprise knowledge graph construction, a simpler domain knowledge graph built on the previous work, oriented towards the enterprise domain, through identified entities and predefined relationships. Named entity recognition of listed company announcements is first implemented in Chapter 2, identifying a large number of names, places, companies, and dates. Then, in the

third chapter, extraction of business-to-business and person-to-business relationships is implemented, and a large number of relationships between entities are obtained. Finally, the entities and relationships are formed into a triad, which is used to construct the enterprise knowledge graph. We are content with first enterprise knowledge graph construction, then knowledge graph relationship evolution based on time constraints, and finally system application implementation.

4.1. Enterprise Knowledge Graph Construction. The basic unit of this web-like knowledge base is the entity-relationship-entity. Both entities and relations can have attributes. Knowledge graphs are useful for information retrieval, using the use of graphically structured data as retrieval output allows the user to visualise the desired results. Currently knowledge graphs are divided into open knowledge graphs and domain knowledge graphs, and our work is to build a relatively simple. We can add more knowledge in the future to build a larger and richer knowledge graph. The process of building a knowledge map is as follows. The process of building a knowledge graph is roughly illustrated in Figure 2.

From Figure 2, the knowledge graph construction process, this paper uses unstructured data, followed by relationship extraction and entity extraction, and finally data layer construction to form the knowledge graph.

The preparatory work was divided into three steps. The first is data preparation. In Chapter 2, a large number of names of people, places, enterprises, and dates have been identified through the identification of named entities in the announcements of listed companies. In Chapter 3, a total of 11 relationships of practical significance have been extracted through the extraction of relationships between companies and companies and between people and companies. The final triad of <person, relationship, enterprise> and <enterprise, relationship, enterprise> is used as the base data to prepare for the construction of the enterprise knowledge graph.

The next step is the structure of the knowledge graph. Figure 3 shows the attribute graph model, where a graph records its nodes and relationships, which can be related to each other, and the corresponding nodes and relationships can have their own attributes. The knowledge graph is a cumulative organisation of information by means of a graph structure, which is managed in a more efficient way, and the knowledge graph has functions such as continuous updating and relational complementary reasoning. Notes recorded in our diagram are the personal and business name entities, and the relationships recorded are the relationships defined in Chapter 3; the relationships contain temporal attributes.

Finally, the choice of tool differs from one tool to another in the way data are stored and presented. First of all, for the huge number of nodes and relationships, graph databases are undoubtedly the best choice and relational databases do not perform very well on data with graph structure, when there are more and more nodes and relationships, the storage and query efficiency of relational databases will be very weak. Graph data belong to the category of NoSQL.

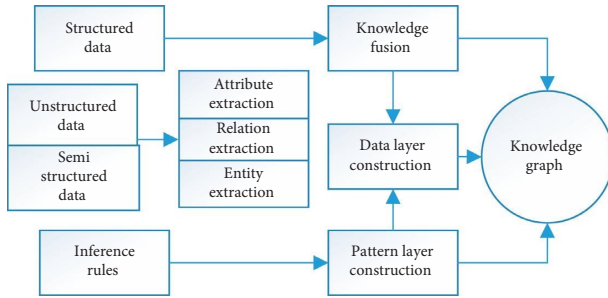


FIGURE 2: Knowledge graph construction process.

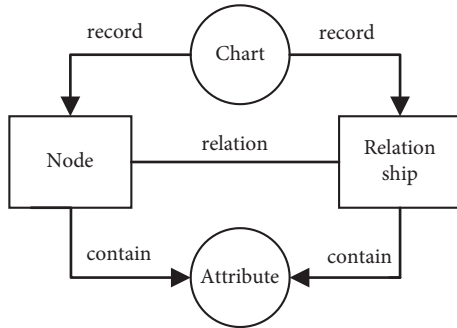


FIGURE 3: Attribute map model.

4.2. The Evolution of Relationships under Time Constraints. It is the inclusion of temporal knowledge information in the knowledge graph that makes it possible to use the knowledge graph to implement simple relationship evolution tasks. Relationship evolution can be seen as an upper-level task of the knowledge graph. For example, there are also upper-level tasks such as relationship complementation and application-level tasks such as knowledge retrieval and antifraud research [27].

We define relational evolution as the existence of certain relations that, subject to temporal constraints, are pushing away, and their relationships have changed or disappeared in the present. As shown in Table 1, it can be known that it is effective to mine known evolution rules for the enterprise knowledge graph.

Table 1 lists the rules for relational evolution from the current relationships, and if richer relationships are added to the knowledge graph, more relevant relationships can be evolved. The relationship evolution is divided into two graphs: AllGraph and NowGraph; before the relationship evolution, all existing entity relationships are stored in AllGraph, and the entity relationships that have evolved are stored in NowGraph.

If a new triple needs to be inserted into the knowledge graph, the triple is first created directly in AllGraph and then in NowGraph of the relationship evolution rules. Figure 4 shows the evolution of the changed relationships.

4.3. Enterprise Knowledge Mapping System Application Implementation. The FTZ system adopts a B/S architecture, with the platform being accessed through a web browser and all data analysis and data queries placed on the server side.

The platform is developed using a front and back-end separation approach [12].

Figure 5 shows the main flow of the prototype system in this paper, in which the analysis building module has mainly three parts: named entity identification, relationship extraction, and knowledge graph construction. In addition, data crawlers are included for the acquisition of listed company announcements, and visualisation pages are created for the display of enterprise knowledge graphs.

5. Case Studies

As shown in Figure 6, through the clustering effect of different enterprises, it is known that by constructing and applying the enterprise knowledge map to the FTZ project, the FTZ management also explains the use, mentioning that the method realizes the service application of the enterprise relationship map for the FTZ by sensing and mastering the data resources that are of value and significance to the FTZ, the method overcomes the administrative internal platform that is not rich in static data, and the method overcomes the difficulties of the internal platform of administration which is not rich in static data and realizes the application practice of web-based Big Data in the work of government regulation and services.

The density of the wood forest products trade network of RCEP member countries from 2001 to 2018 has generally fluctuated slightly but still shows a gradual increase (Figure 7), which indicates that the trade links between member countries in the wood forest products trade network of RCEP member countries are relatively close, the scale of trade is expanding, and the trade network is becoming increasingly complex. The uneven distribution of forest resources in the RCEP member countries makes it impossible to achieve optimal allocation of wood forest products within the RCEP member countries, so as the demand for wood forest products grows, trade in wood forest products between RCEP member countries is becoming more frequent.

As shown in Figure 8, a social network approach was used to analyse the product trade network pattern among member countries based on the product import and export trade data of RCEP member countries from 2001 to 2018. The study shows that the product trade links between member countries are relatively strong and the trade relations are becoming increasingly complex; the product trade network between member countries is developing towards an optimal network pattern, and the product trade exchanges are becoming more balanced [14].

Taken advantage of Shanghai's natural ecological environment to create a garden style office environment for the enterprises in the park, on the other hand, it has built a high-quality industrial ecological environment and set up an enterprise service supermarket to provide "one-stop" services for industry and commerce, taxation, and social security. Figure 9 shows the impact of different parameters on the scheme of this paper. China's product trade position has replaced Japan's core position in the product trade network as the only core country.

TABLE 1: Rules for the evolution of relationships.

Relationship	Evolution rules
Increase holdings	Shareholding increase - > equity participation
Reduce one's shares in a listed company	Reduction - > participation
Participation - > withdrawal	Currently not relevant
Change	Link the old node relationship information to the new node after the change

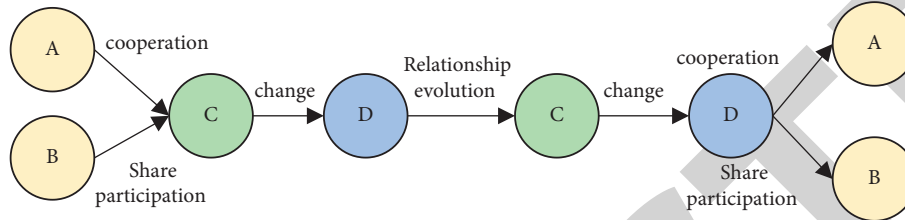


FIGURE 4: Evolution of change relationships.

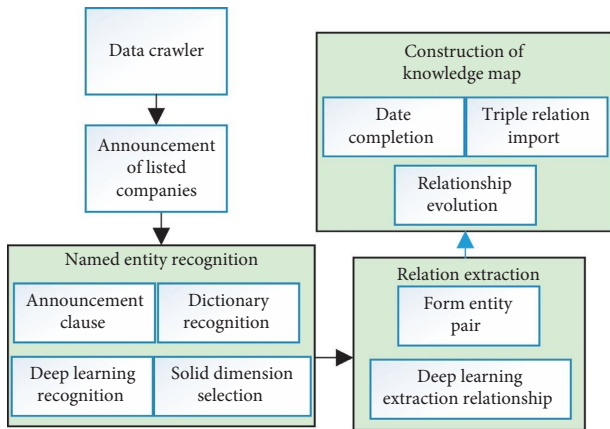


FIGURE 5: Overall system framework.

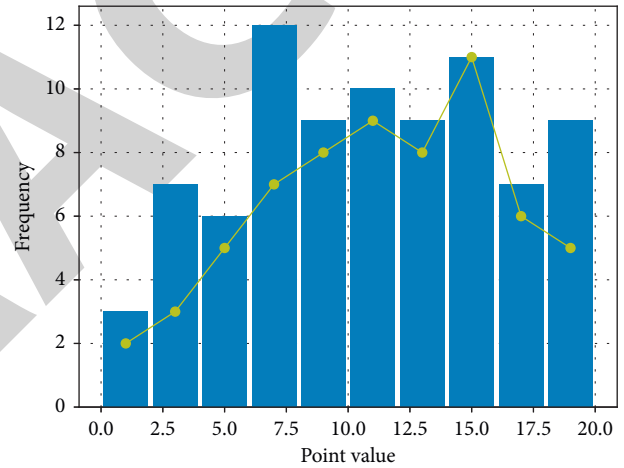
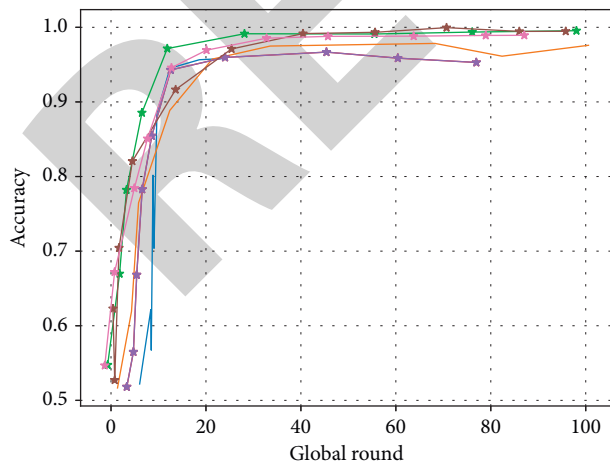


FIGURE 7: FTA sales volatility.



- FTZ
- FedProx
- FLDA
- C-FLA
- FLA
- OUR

FIGURE 6: Clustering effect of knowledge nodes for different enterprises.

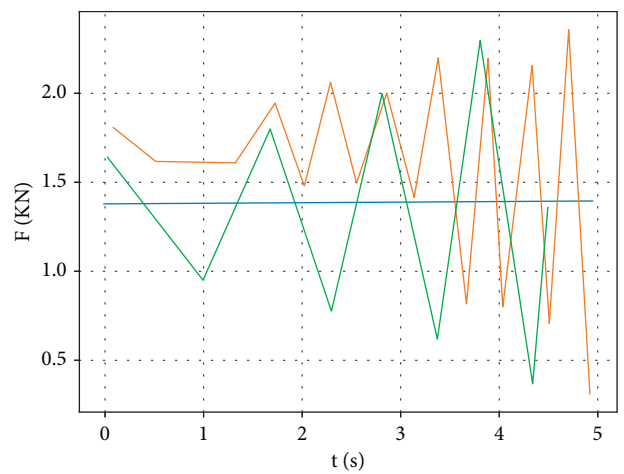


FIGURE 8: Efficiency of business registration by region.

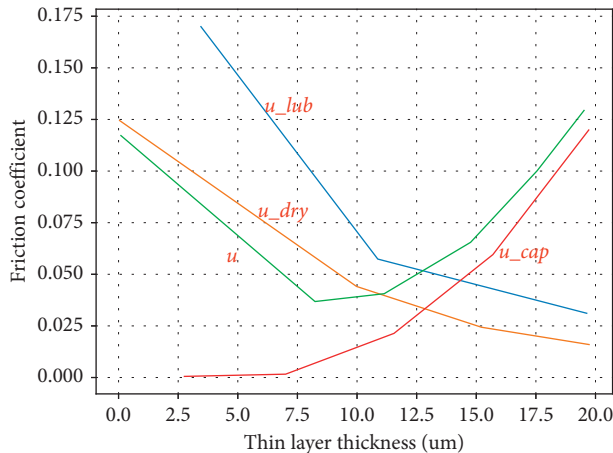


FIGURE 9: Effect of different parameters on the method of this paper.

6. Conclusions

We present the process of building the enterprise knowledge graph and its final application to the ZFT project. Firstly, we did some preparatory work for the construction of the enterprise knowledge graph, firstly data preparation and then the construction of the knowledge graph based on Neo4j using a suitable knowledge graph structure. On top of the constructed enterprise knowledge graph, we performed a relational evolution. Finally, the application of the enterprise knowledge graph was implemented. Firstly, we gave a brief introduction to the FTZ project and introduced the development techniques related to the platform, then the knowledge graph constructed by our method was deployed to the project, and finally the knowledge graph was presented using search statements, which was well evaluated by the FTZ management through practical use.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] T. Elwert, D. Goldmann, F. Römer et al., "Current developments and challenges in the recycling of key components of (hybrid) electric vehicles," *Recycling*, vol. 1, no. 1, pp. 25–60, 2015.
- [2] D. Ahlstrom, M. N. Young, and A. Nair, "Navigating China's feudal governance structures: some guidelines for foreign enterprises[J]," *Sam Advanced Management Journal*, vol. 68, no. 1, pp. 4–13, 2003.
- [3] S. Hua, N. Hu, and H. H. Yang, "Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes[J]," *Clinical Cancer Research An Official Journal of the American Association for Cancer Research*, vol. 17, no. 9, pp. 2955–2966, 2011.
- [4] G. E. Curtis, S. L. Elan, R. A. Hudson, and N. A. Kollars, "Transnational activities of Chinese crime organizations," *Trends in Organized Crime*, vol. 7, no. 3, pp. 19–57, 2002.
- [5] S. U. Hao and C. Yang, "Apec summit:building a reasonable institutional framework for asia-pacific free trade zone[j]," *Contemporary World*, vol. 2017, no. 01, pp. 44–47, 2016.
- [6] J. Nkem, F. B. Kalame, M. Idinoba, O. A. Somorin, O. Ndoye, and A. Awono, "Shaping forest safety nets with markets: adaptation to climate change under changing roles of tropical forests in Congo Basin," *Environmental Science & Policy*, vol. 13, no. 6, pp. 498–508, 2010.
- [7] H. Y. Su, "Apec summit:building a reasonable institutional framework for asia-pacific free trade zone[J]," *Procedia - Social and Behavioral Sciences*, vol. 2021, no. 2017-1, pp. 42–45, 2016.
- [8] D. Gueorguiev and E. Malesky, "Foreign investment and bribery: a firm-level analysis of corruption in Vietnam," *Journal of Asian Economics*, vol. 23, no. 2, pp. 111–129, 2012.
- [9] M. Hebblewhite, D. G. Miquelle, H. Robinson et al., "Including biotic interactions with ungulate prey and humans improves habitat conservation modeling for endangered Amur tigers in the Russian Far East," *Biological Conservation*, vol. 178, pp. 50–64, 2014.
- [10] K. G. Ho, "Winning e-commerce strategy for small and medium enterprises (SMEs) to meet customers' needs in Hong Kong and greater China region[J]," *Agu Fall Meeting Abstracts*, vol. 12, no. 1, pp. 93–106, 2010.
- [11] S. Norsukhairin, H. Juahir, A. Azid, S. M. Sharif, and M. T. Latif, "Spatial and temporal air quality pattern recognition using environmetric techniques: a case study in Malaysia[J]," *Environmental science-processes & impacts*, vol. 15, no. 9, pp. 1717–1728, 2013.
- [12] O. Yi, "Comment on "regionalism in the asia-pacific" and U. S. Interests[J]," *Nanzan Review of American Studies a Journal of center for American Studies Nanzan university*, vol. 30, pp. 175–183, 2008.
- [13] C. Nick, "Damming Afghanistan: modernization in a buffer state[J]," *Journal of American History*, vol. 2002, no. 2, pp. 512–537, 2002.
- [14] P. Ward, "Can more information Be shared in China?[J]," *China's Foreign Trade*, vol. 2011, no. 17, p. 94, 2011.
- [15] B. Benjamin, A. Paul, B. Jürgen, and C. Jason, "Mapping local climate zones for a worldwide database of the form and function of cities[J]," *International Journal of Geo-Information*, vol. 4, no. 1, pp. 199–219, 2015.
- [16] L. I. Jing, "Web information extraction[J]," *Computer Science*, vol. 9, no. 5, pp. 823–827, 2003.
- [17] J. Blazevic, "Parallel worlds[J]," *D-lib Magazine*, vol. 24, no. 3, pp. 120–121, 2005.
- [18] X. M. Chen, "China's steady progress in building a harmonious world[J]," *The Chinese Journal of Global Governance*, vol. 2021, no. 2013-1, pp. 10–13, 2021.
- [19] S. K. Kil, "Establishment of the logistics hub in northeast asia on the basis of LME warehouses[J]," *Ocean Policy Research*, vol. 18, pp. 1–32, 2003.
- [20] H. Hengsdijk and M. K. van Ittersum, "Formalizing agro-ecological engineering for future-oriented land use studies," *European Journal of Agronomy*, vol. 19, no. 4, pp. 549–562, 2003.
- [21] G. Fagiolo and A. Roventini, "Macroeconomic policy in DSGE and agent-based models[J]," *Social Science Electronic Publishing*, vol. 124, no. 5, pp. 67–116, 2016.

- [22] H. Su, Z. H. Jiang, and H. W. Wu, "Building knowledge map for product development[J]," *Journal of Shanghai Jiaotong University*, vol. 39, no. 12, pp. 2034–2039, 2005.
- [23] P. Subacchi, "The people's money: how China is building a global currency[J]," *China*, vol. 2017, no. 2, p. 64, 2017.
- [24] X. Liu, "CCPIT proactively participates in the construction of pilot free trade zones[J]," *China's Foreign Trade*, vol. 565, no. 01, pp. 37–39, 2018.
- [25] M. H. Ditmore and D. Allman, "'Who is Helsinki?' Sex workers advise improving communication for good participatory practice in clinical trials," *Health Education Research*, vol. 26, no. 3, pp. 466–475, 2011.
- [26] C. Cao, Y. Tang, D. Huang, W. Gan, and C. Zhang, "IIBE: an improved identity-based encryption algorithm for wsn security," *Security and Communication Networks*, vol. 1, pp. 1–8, Article ID 8527068, 2021.
- [27] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," *Bio-inspired Computing Theories and Applications*, Springer, vol. 2016, pp. 513–517, Singapore, 2016.