Please share the highlighted papers in my Email- [rcs23102@iiitl.ac.in](mailto:rcs23102@iiitl.ac.in)

| **Transformer-Based Named Entity Recognition in**  **Construction Supply Chain Risk Management**  **in Australia** | In the Australian construction industry, effective supply chain risk management (SCRM) is  critical due to its complex networks and susceptibility to various risks. This study explores the application  of transformer models like BERT, RoBERTa, DistilBERT, ALBERT, and ELECTRA for Named Entity  Recognition (NER) in this context. Utilizing these models, we analyzed news articles to identify and classify  entities related to supply chain risks, providing insights into the vulnerabilities within this sector. Among  the evaluated models, RoBERTa achieved the highest average F1 score of 0.8580, demonstrating its superior  balance in precision and recall for NER in the Australian construction supply chain context. Our findings  highlight the potential of NLP-driven solutions to revolutionize SCRM, particularly in geo-specific settings. | The use of advanced language models, like BERT and  GPT-3, in NER has become increasingly prevalent across  various industries. From healthcare to finance, legal, and  construction, businesses are leveraging these sophisticated  models to accurately identify and categorize named entities  within large volumes of text. These models have the  remarkable ability to autonomously detect complex patterns  and relationships between words without the need for labour-  intensive feature engineering. This capability allows for a  nuanced understanding of data, enabling critical insights  extraction, better decision-making, regulatory compliance,  and improved customer experiences. Additionally, advance-  ments in transfer learning and the development of domain-  specific pre-trained models have further accelerated the  effectiveness and adoption of NER across diverse industries.  In today’s data-driven ecosystem, NER has become an  indispensable tool [23], [24].  Word2Vec (W2V) has revolutionized semantic vector  spaces in NLP, building on earlier foundations [25]. It intro-  duces word embeddings through two methods: Continuous  Bag-of-Words (CBOW) and Skip-Gram (SG), both sharing  a neural network structure but differing in input-output man-  agement [26], [27]. Evolving beyond basic word embeddings,  multi-sense and contextualized embeddings like Elmo, Bert,  and Xlnet have emerged, focusing on enriched semantic  understanding [28]. W2V bridges the gap between count-  based models and neural networks, enhancing semantic  exploration and text analytics in deep learning, thus playing  a pivotal role in the evolution of pre-trained language  models [29]. Reference [30] created a NER methodology to identify  Chinese medicine and disease names in conversations  between humans and machines. They evaluated various  models, and the combination of RoBERTa with biLSTM  and CRF performed the best. Using a corpus obtained  through web crawling, this model achieved an impressive  Precision, Recall, and F1-score of 0.96. These findings  highlight its potential for enhancing medication reminders  in dialogue systems. Reference [31] developed a Chi-  nese NER model called BBIEC specifically for analysing  COVID-19 epidemiological data. This model effectively  processes unlabelled data at the character level, extracting  global and local features using pre-trained BERT, BiLSTM,  and IDCNN techniques. The BBIEC model outperforms  traditional models when it comes to recognizing entities  that are crucial for analysing the transmission routes and  sources of the epidemic. Reference [32] proposed a BERT-  Transformer-CRF based service recommendation method  (BTC-SR) for enhanced chronic disease management, which  initially employs a BERT-Transformer-CRF model to iden-  tify named entities in disease text data, extracts entity  relationships, and integrates user implicit representation  to deliver personalized service recommendations, demon-  strating improved entity recognition with an F1 score of  60.15 on the CMeEE dataset and paving the way for  more precise service recommendations for chronic disease  patients.  Reference [33] introduced a deep learning-based Mineral  Named Entity Recognition (MNER) model, utilizing BERT  for mineral text word embeddings and enhancing sequence  labelling accuracy by integrating the CRF algorithm’s  transfer matrix. Furthermore, [34] introduced a multi-task  model called BERT-BiLSTM-AM-CRF. The model utilizes  BERT for dynamic word vector extraction and then refines it  through a BiLSTM module. After incorporating an attention  mechanism network, the output is passed into a CRF layer  for decoding. The authors tested the model on two Chinese  datasets and observed significant improvements in F1 score  compared to previous single-task models, with increases of  0.55% in MASR dataset and 3.41% in People’s Daily dataset  respectively. Reference [35] explored the NER task in Telugu  language using various embeddings such as Word2Vec,  Glove, FastText, Contextual String embedding, and BERT.  Remarkably, when combining BERT embeddings with  handcrafted features, the results outperformed other models  significantly. The achieved F1-Score was an impressive  96.32%. Reference [36] introduced Wojood, a unique corpus  specifically designed for Arabic nested NER. This corpus  comprises approximately 550K tokens of Modern Standard  Arabic and dialect, each manually annotated with 21 different  entity types. Unlike traditional flat annotations, Wojood  includes around 75K nested entities, accounting for about  22.5% of the total annotations. The accuracy and reliability  of this corpus are evident in its substantial interannotator  agreement, with a Cohen’s Kappa score of 0.979 and an  F1 score of 0.976. Furthermore, to address the limitations of traditional methods for named entity recognition in  the context of agricultural pest information extraction,  [37] proposed a PBERT-BiLSTM-CRF model. This model  leverages pre-trained BERT to resolve ambiguity, BiLSTM  to capture long-distance dependencies, and CRF for optimal  sequence annotation. The results demonstrate significant  improvements in precision, recall, and an impressive F1 score  of 90.24% compared to other models. NAMED ENTITY RECOGNITION IN CONSTRUCTION  INDUSTRY  Named entity recognition in construction has received some  attention in academic literature, although the available  published research in this field is relatively limited. While  several studies have been conducted on this topic, the  quantity of publications compared to other areas of natural  language processing and construction is modest. In the  realm of CSCRM in Australia, the significance of local and  international news cannot be overstated.  The constantly changing geopolitical, environmental, and  economic scenarios greatly impact construction supply  chains. For example, the recent disruptions caused by the  COVID-19 pandemic had a profound effect on the China-  Australia construction supply chain. This highlighted the  urgent need for timely and accurate information to effectively  manage and mitigate risks [38]. The construction sector in  Australia is currently facing increased supply chain risks.  These risks have been amplified by the growing number  of suppliers, complex work streams, stringent compliance  requirements, and difficulties in finding eligible parties. It is  important to note that disruptions in global supply chains,  particularly those originating from regions like China, have  resulted in project delays. This emphasizes the significance  of international news for predicting and managing such  disruptions. The lack of transparency in supply chain  risk among Australian construction firms emphasizes the  need for a well-informed and data-driven approach to risk  management. By utilizing NER technologies, particularly in  the context of geological news texts, automation can play a  vital role in extracting relevant information from local and  international news sources. This enhancement significantly  improves the accuracy and timeliness of risk assessments and  mitigating actions within the Australian construction supply  chain domain.  However, the field of geological news texts is rapidly  expanding, offering a wealth of valuable information.  Accurately extracting this information can greatly enhance  geological survey efforts. However, traditional manual  extraction methods are inefficient and time-consuming,  leading to lower accuracy. As the volume of geological  news text data increases, these challenges become even more  pronounced. It is crucial to transition towards automated  extraction paradigms to address this complexity. Automating  the extraction of geological news entities goes beyond  just a procedural evolution; it represents a fundamental leap towards the creation of comprehensive geological  knowledge graphs. These knowledge graphs can serve as  structured repositories, facilitating the retrieval and analysis  of geological information and propelling advancements in the  field of geological surveys.  BERT, however, is a major breakthrough in the field of  deep language understanding. Its architecture, which utilizes  the powerful Transformer model, particularly its encoder  component, has revolutionized our ability to comprehend nat-  ural language. BERT’s pre-training phase involves analysing  an enormous corpus of books and Wikipedia articles, allow-  ing it to grasp the complex semantics present in textual data.  The core essence of BERT lies within the encoder section  of the Transformer model–an innovative design introduced  by [39],–which has received widespread acclaim for its  efficient parallelization of computations, greatly improv-  ing computational efficiency. The recent advancements in  machine learning and NLP have significantly improved  the challenges associated with manual data extraction. One  notable breakthrough is the emergence of transformer-based  models like BERT, which has paved the way for automating  the extraction process. For example, a study introduced a  method called Geological News Named Entity Recognition  (GNNER) that utilizes the BERT language model to effec-  tively extract and leverage geological data [40]. Moreover,  other scholarly endeavours have demonstrated automated  techniques for extracting spatiotemporal and semantic infor-  mation from geological documents. These techniques are  crucial for tasks such as data mining, knowledge discovery,  and constructing knowledge graphs [41], [42]. The narrative  above explains the importance and modern approaches used  in automating the extraction of geological news information.  This automation not only enhances the efficiency and  accuracy of retrieving information, but it also forms a vital  foundation for building comprehensive geological knowledge  graphs. Table 1 compares the recent literature on NER in  construction industry with current study considering their  aims, models and their dataset used. |
| --- | --- | --- |

[23] C. Berragan, A. Singleton, A. Calafiore, and J. Morley, ‘‘Transformer based named entity recognition for place name extraction from unstructured text,’’ Int. J. Geographical Inf. Sci., vol. 37, no. 4, pp. 747–766,Apr. 2023.

[24] C.-M. Tsai, ‘‘Stylometric fake news detection based on natural language

processing using named entity recognition: In-domain and cross-domain

analysis,’’ Electronics, vol. 12, no. 17, p. 3676, Aug. 2023.

[25]Y. Bengio, R. Ducharme, and P. Vincent, ‘‘A neural probabilistic language

model,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 13, 2000, pp. 1–12.

[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean, ‘‘Efficient estimation of

word representations in vector space,’’ 2013, arXiv:1301.3781.

[27] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ‘‘Dis-

tributed representations of words and phrases and their compositionality,’’

in Proc. Adv. Neural Inf. Process. Syst., vol. 26, 2013, pp. 1–13.

[28] S. J. Johnson, M. R. Murty, and I. Navakanth, ‘‘A detailed review on word

embedding techniques with emphasis on word2vec,’’ Multimedia Tools

Appl., vol. 2023, pp. 1–29, Oct. 2023.

[29] F. Almeida and G. Xexéo, ‘‘Word embeddings: A survey,’’ 2019,

arXiv:1901.09069.

[30] T.-H. Yang, M. Pleva, D. Hládek, and M.-H. Su, ‘‘BERT-based Chinese

medicine named entity recognition model applied to medication reminder

dialogue system,’’ in Proc. 13th Int. Symp. Chin. Spoken Lang. Process.

(ISCSLP), Dec. 2022, pp. 374–378.

[31] C. Yang, L. Sheng, Z. Wei, and W. Wang, ‘‘Chinese named entity

recognition of epidemiological investigation of information on COVID-

19 based on BERT,’’ IEEE Access, vol. 10, pp. 104156–104168, 2022.

[32] D. Chen, C. Liu, and Z. Zhao, ‘‘Named entity recognition service

of BERT-transformer-CRF based on multi-feature fusion for chronic

disease management,’’ in Proc. Int. Conf. Service Sci. Cham, Switzerland:

Springer, 2023, pp. 166–178.

[33] Y. Yu, Y. Wang, J. Mu, W. Li, S. Jiao, Z. Wang, P. Lv, and Y. Zhu,

‘‘Chinese mineral named entity recognition based on BERT model,’’

Expert Syst. Appl., vol. 206, Nov. 2022, Art. no. 117727.

[34] X. Tang, Y. Huang, M. Xia, and C. Long, ‘‘A multi-task BERT-

BiLSTM-AM-CRF strategy for Chinese named entity recognition,’’

Neural Process. Lett., vol. 55, no. 2, pp. 1209–1229, Apr. 2023.

[35] S. Gorla, S. S. Tangeda, L. B. M. Neti, and A. Malapati, ‘‘Telugu named

entity recognition using bert,’’ Int. J. Data Sci. Anal., vol. 14, no. 2,

pp. 127–140, Aug. 2022.

[36] M. Jarrar, M. Khalilia, and S. Ghanem, ‘‘Wojood: Nested Arabic named

entity corpus and recognition using BERT,’’ in Proc. 13th Lang. Resour.

Eval. Conf., 2022, pp. 3626–3636.

[37] Z. Lun and Z. Hui, ‘‘Research on agricultural named entity recognition

based on pre train BERT,’’ Academic J. Eng. Technol. Sci., vol. 5, no. 4,

pp. 34–42, 2022.

[38] C. V. Ndukwe, J. Liu, and T. K. Chan, ‘‘Impact of COVID-19 on

the China–Australia construction supply chain,’’ in Proc. 25th Int.

Symp. Advancement Construct. Manage. Real Estate. Cham, Switzerland:

Springer, 2021, pp. 1275–1291.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,

Ł. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ in Proc. Adv.

Neural Inf. Process. Syst., vol. 30, 2017, pp. 1–11.

[40] C. Huang, Y. Wang, Y. Yu, Y. Hao, Y. Liu, and X. Zhao, ‘‘Chinese named

entity recognition of geological news based on BERT model,’’ Appl. Sci.,

vol. 12, no. 15, p. 7708, Jul. 2022.

[41] Q. Qiu, Z. Xie, L. Wu, and L. Tao, ‘‘Automatic spatiotemporal and

semantic information extraction from unstructured geoscience reportsp

using text mining techniques,’’ Earth Sci. Informat., vol. 13, no. 4,

pp. 1393–1410, Dec. 2020.

[42] X. Lv, Z. Xie, D. Xu, X. Jin, K. Ma, L. Tao, Q. Qiu, and Y. Pan, ‘‘Chinese

named entity recognition in the geoscience domain based on BERT,’’

Earth Space Sci., vol. 9, no. 3, Mar. 2022, Art. no. e2021EA002166.