



Review article

Evolution and emerging trends of named entity recognition: Bibliometric analysis from 2000 to 2023

Jun Yang^a, Taihua Zhang^{a,b}, Chieh-Yuan Tsai^c, Yao Lu^{a,b}, Ligu Yao^{a,b,*}

^a School of Mechanical and Electrical Engineering, Guizhou Normal University, Guiyang, Guizhou, 550025, China

^b Technical Engineering Center of Manufacturing Service and Knowledge Engineering, Guizhou Normal University, Guiyang, Guizhou, 550025, China

^c Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan, 32003, Taiwan

ARTICLE INFO

Keywords:

Named entity recognition
CiteSpace
Natural language processing
Bibliometrics

ABSTRACT

Identifying valuable information within the extensive texts documented in natural language presents a significant challenge in various disciplines. Named Entity Recognition (NER), as one of the critical technologies in text data processing and mining, has become a current research hotspot. To accurately and objectively review the progress in NER, this paper employs bibliometric methods. It analyzes 1300 documents related to NER obtained from the Web of Science database using CiteSpace software. Firstly, statistical analysis is performed on the literature and journals that were obtained to explore the distribution characteristics of the literature. Secondly, the core authors in the field of NER, the development of the technology in different countries, and the leading institutions are explored by analyzing the number of publications and the cooperation network graph. Finally, explore the research frontiers, development tracks, research hotspots, and other information in this field from a scientific point of view, and further discuss the five research frontiers and seven research hotspots in depth. This paper explores the progress of NER research from both macro and micro perspectives. It aims to assist researchers in quickly grasping relevant information and offers constructive ideas and suggestions to promote the development of NER.

1. Introduction

With the arrival of the information era, text data in various fields has grown exponentially. A lot of valuable professional information is covered in semi-structured or unstructured text recorded in natural language. How to mine these from massive text data information has become a research hotspot in various fields. It is usually time-consuming and error-prone to manually extract information from these data, so the method of extracting information from texts using artificial intelligence technology comes into being.

The concept of named entity (NE) was first used in the Message Understanding Conference - 6 (MUC-6) [1], in which the main concerned entity categories are people, organizations, places, time expressions, etc. (general field). In a specific subject field, NE refers to the object of concern. For example, in biology, it refers to proteins, genes, diseases [2,3], and so on. In chemistry, it refers to compounds, solvents [4,5], and so on. NER is a crucial and essential task in text mining, which aims to identify the types and boundaries of NE. For the label sequence $S = \langle w_1, w_2, \dots, w_n \rangle$ of a given text, one or more triplet lists $\langle I_s, I_e, t \rangle$ are obtained after the NER model, each triplet contains one entity information, in the triplet, I_s shows the beginning index position of the entity, I_e shows the

* Corresponding author.

E-mail addresses: juny@gznu.edu.cn (J. Yang), zhangth542@gznu.edu.cn (T. Zhang), cytsai@saturn.yzu.edu.tw (C.-Y. Tsai), yao.lu@gznu.edu.cn (Y. Lu), lgyao@gznu.edu.cn (L. Yao).

<https://doi.org/10.1016/j.heliyon.2024.e30053>

Received 3 January 2024; Received in revised form 16 April 2024; Accepted 18 April 2024

Available online 22 April 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

terminate index position of the entity, and t indicates the entity type. Its principal structure is shown in Fig. 1. Given a paragraph of text “Xiao Li Lives in Beijing, the capital of China.” three triples are obtained by the NER model, $\langle w_1, w_2, \text{Person} \rangle$, $\langle w_5, w_5, \text{Location} \rangle$, $\langle w_{10}, w_{10}, \text{Organization} \rangle$, where w_1 shows the beginning index position of entity “Person”, w_2 means the terminate index position of the entity “Person”, and “Person” is the entity type. NER not only plays a crucial role in the field of text mining but also in natural language processing (NLP) such as information retrieval [6], automatic text summarization [7], question answering system [8], machine translation [9], knowledge graph [10], etc. applications also play an essential role.

Early NER methods can be divided into two categories: rule-based methods [11,12] and machine learning-based methods [13,14]. The rule-based method means that experts manually create rule templates, and then the entity is obtained by matching according to the rule template or lexicon. Although this method has high recognition accuracy, it has high labor costs and poor generalization ability. Some well-known rule-based NER systems include Lasie-II [15], NetOw1 [16], Facile [17], and FASTUS [18], etc. These systems are mainly based on manually customized semantic and grammatical rules to recognize entities. Machine learning-based methods transform the NER task into a sequence labeling task, which uses a large-scale annotated corpus to train the annotation model to tag each token in the text through the trained model. Then, the automatically tagged sequence is decoded according to the tagging scheme and integrated into the NE composed of several characters in the text [19]. Machine learning algorithms can be divided into three categories according to whether the training dataset is labeled: supervised learning, semi-supervised learning, and unsupervised learning. Common machine algorithms include the Hidden Markov Model (HMM) [20], Maximum Entropy Model (MENE) [21], Support Vector Machine (SVM) [22], Decision Tree (DT) [23] and Conditional Random Field (CRF) [24], etc. In recent years, deep learning has proved to be an effective strategy to extract feature representations directly from text data, which has made breakthroughs in the field of NER. Compared with methods based on statistical learning, deep learning makes it easier to discover hidden features due to the characteristics of multi-layer nonlinearity [25].

Although NER has been developing for decades, there are few reviews in this field. In 2013, Marrero et al. [26] conducted an in-depth discussion on the application, evaluation methods, and different definitions of named entities of NER, with special emphasis on the research mainstream of machine learning-based and rule-based NER technology at that time. With the rise of deep learning technology, the NER field has experienced significant changes. Goyal et al. [27] provided a comprehensive overview of the development status of NER and classification technology and explored diverse technical paths from rule-based methods to unsupervised learning. Nasar et al. [28] conducted an extensive review of methods for NER and relationship extraction, highlighting the advantages of hybrid and joint models based on deep learning. Their research revealed the significant contribution of deep learning technology in improving recognition accuracy and processing complex entity relationships. Li et al. [29] focused on introducing the NER method based on deep learning by subdividing NER technology into a distributed representation of the input, context encoder, and label decoder. They not only demonstrate the impact of deep learning techniques in standardizing model structures but also systematically classify existing work. Previous research has deeply explored various aspects of NER technology, contributing important insights to the development of the field. However, most of them focus on evaluating specific technologies or methods, and this focused perspective rarely touches on the macro development trends of NER research. Likewise, there is a relative lack of comprehensive assessments on the evolution of research hotspots and cutting-edge technologies.

Bibliometrics can quantitatively analyze the advanced trends and research hotspots of the field based on published literature [30], and it is also an objective and scientific analysis method. Through this method, researchers can explore important topics and their interrelationships in the research field and deeply understand the process of knowledge sharing and diffusion, thus providing valuable insights for future research directions and policy formulation. For example, Yu and Pan [31] applied bibliometric methods to deeply

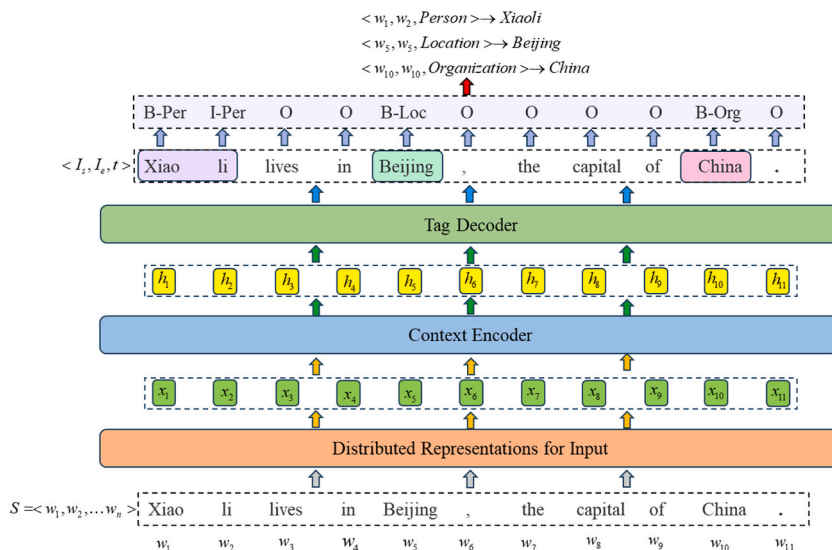


Fig. 1. structural schematic diagram of NER.

explore and analyze the knowledge development process in the research field of Technique for Order Preference by Similarity to an Ideal Solution. Through a comprehensive survey of key literature transmission paths in citation networks, this paper reveals the knowledge diffusion model and its development trajectory over time in this field. Furthermore, it delves into the intricate knowledge structure and specialized research topics within the research community of this field. Yu et al. [32] analyzed literature related to intuitionistic fuzzy set theory through bibliometrics, which provided a macro perspective on the evolution of research in this field and vividly demonstrated the evolution of topics in this field. The knowledge diffusion path in this field was explored through the main path analysis of global and critical paths. These studies demonstrate the powerful application capabilities of bibliometrics in the field of scientific research and highlight its value as a scientific research tool. CiteSpace is a bibliometrics visualization software based on Java language [33], providing powerful tools to objectively reveal development trends and research hotspots in the scientific field. The software can analyze and visualize citation relationships, co-citation networks, and keyword co-occurrence networks in documents, thereby directly displaying mainstream research directions and key issues in the field [34]. At the same time, through detailed visual display, CiteSpace can depict the evolution of the knowledge structure and the interaction of the research community in the NER field, providing a basis for further research. This is particularly important for NER, a multifaceted and rapidly developing field because valuable information and trends from a large amount of academic literature need to be extracted. In this context, this study is different from the previous analysis that mainly focused on specific technologies or methods and adopts the method of bibliometrics to analyze relevant documents in the Web of Science database. It not only discusses the overall trend of NER research, key research hotspots, and how they evolve over time from a macro perspective but also focuses on the research frontiers and related research hotspots in this field through in-depth analysis of relevant literature. Through in-depth mining and visual display of literature data, the broad layout of the research network in the NER field is depicted, including the distribution of leading institutions and countries/regions, providing a clear and objective perspective for research in this field. It aims to help researchers quickly grasp research frontiers and hotspots and provide constructive ideas and suggestions for promoting the development of NER.

The remainder of the paper is organized as follows: Section 2 introduces the data sources and research methodology, and Section 3 analyzes the number of published papers, research directions, and the distribution of journals. Section 4 analyzes the number of articles published by authors, institutions, and countries and their cooperation. Section 5 explores the research frontiers in the field of NER through the analysis of co-cited literature. Section 6 explores the research hotspots of NER by analyzing keywords. Finally, Section 7 summarizes the paper’s results and presents ideas for further development.

2. Data source and methods

2.1. Data source

The data used in this study are obtained from the Science Citation Index Expanded (SCI-Expanded) and Social Sciences Citation Index (SSCI) databases in the Web of Science core database. SCI-Expanded and SSCI contain many authoritative publications and the most extensive data, so the data obtained by choosing this database is convincing enough. In recent years, NER technology has experienced a transformation from initial exploration to rapid development, especially the rise of deep learning technology, which has brought significant progress and changes to the field. With the availability of large amounts of text data, the development of NER technology has been supported by powerful resources. Before 2000, the number of NER research papers was relatively small, and this history has significance in the context of technological development and theory. However, considering the increasing application of deep learning technology in NER research since 2010, we choose the analysis scope starting from 2000, aiming to capture the

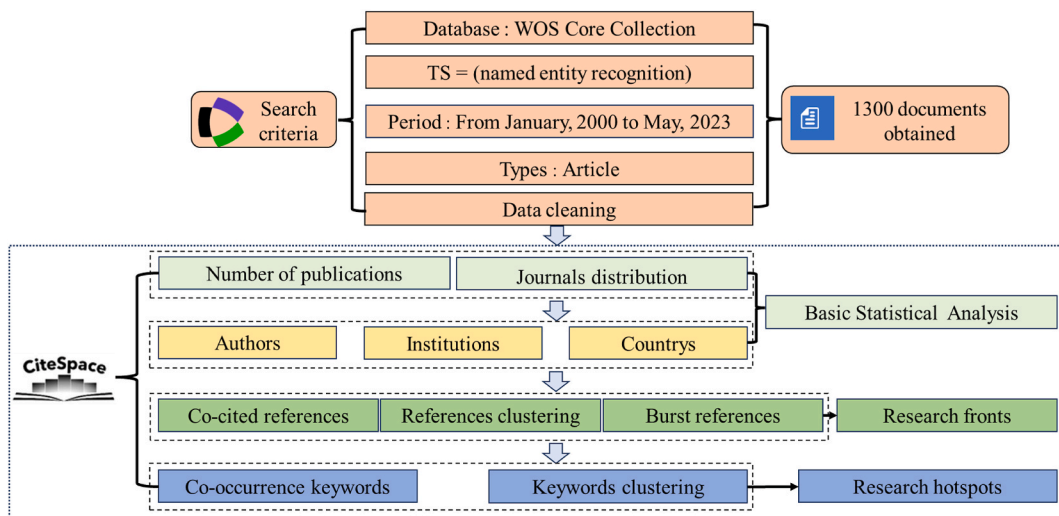


Fig. 2. Research framework diagram of the article.

development process of the transition from traditional methods to deep learning methods. Therefore, we used “TS=Named entity recognition” as the search formula (search time May 2023) and selected the period from 2000 to 2023 to analyze the trajectory of rapid progress and key changes in NER technology during this period. The literature type was chosen as the article. After retrieval, 1913 literature was obtained, and the data was cleaned manually. After removing irrelevant literature, 1300 papers and text research data sources were finally obtained.

2.2. Research methods

The research uses bibliometrics and visualization methods based on relevant literature data by drawing “author cooperation”, “institutional cooperation”, “literature co-occurrence”, “literature clustering”, “citation burst”, “keyword co-occurrence”, and “keyword clustering” network maps, which intuitively and scientifically display the characteristics of documents and the development trend and research frontier of NER. The research framework of this paper is shown in Fig. 2.

3. Literature distribution characteristics

3.1. Analysis of annual publications

To some degree, the yearly publication quantity in this field can indicate its developmental progress. The number of papers published in the 1215 literature records obtained was statistically sorted out. The trend chart of the annual number of papers published in the past 22 years was drawn, as shown in Fig. 3. As the complete count of papers published in 2023 is unavailable, this year’s publications are excluded from the figure. In 2005, the publication numbers reached a short-term peak, possibly related to the ACE Conference 2004. The ACE plan aims to extract the entities mentioned from natural language data and the relationships between these entities and their participation in events [35]. At present, the NER model based on deep learning has become mainstream and has achieved good results, so the NER field has developed rapidly. Price’s Law of Literature Growth states that at the early stage of the birth of an area, the growth of the number of related documents is in an unstable stage; when the field is in a period of rapid development, the number of documents grows exponentially; when the area is in a mature location, the number of documents grows relatively slowly. The mathematical model of the exponential curve is used to fit the number of publications, and the parameters of the curve are obtained by the least square method. The curve-fitting formula for the number of publications is finally accepted, as shown in Eq. (1).

$$y = 11.88 + 0.0069 \times e^{\frac{x-1991.44}{2.88}} \quad (1)$$

where y is the annual number of publications and x is the year. The degree of fit of the curve can be judged by the R^2 (coefficient of determination), and $0 \leq R^2 \leq 1$, the closer the value is to 1, the higher the fitting reliability. The fitting curve $R^2 = 0.977$, indicates that the fitting reliability is high. The red line in Fig. 3 is the fitting curve. It can be seen that the number of publications in the NER field has increased exponentially since 2018, so NER is in a period of rapid development.

3.2. Research directions and journals distribution

Analyzing the research directions of NER can assist in understanding the background knowledge and basic disciplines involved in this technology. Through the function of analyzing the search results provided by the WOS database, the number of papers in each research direction involved in NER is obtained, as shown in Fig. 4. Among them, Computer Science Information Systems had the most significant number of papers, with 471 papers. The second direction was Computer Science and Artificial Intelligence, with 443 papers. They were followed closely by Computer Science Interdisciplinary Applications with 329 papers. Through the analysis of research

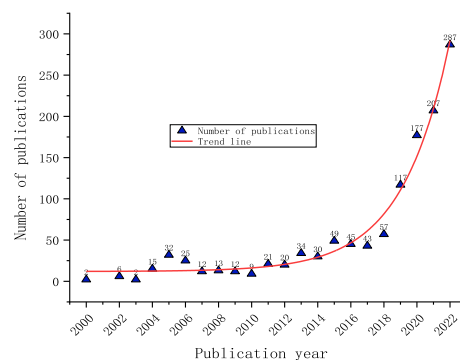


Fig. 3. Annual publication numbers and its trend chart.

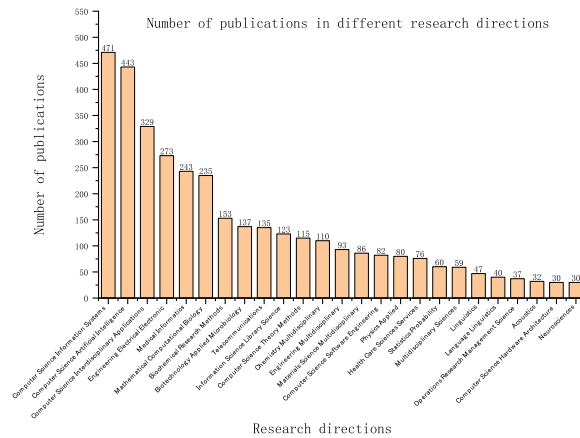


Fig. 4. Number of papers included in different research direction.

directions, it can be known that NER involves artificial intelligence, medicine, computer science, electrical electronics engineering, biology, biochemistry, and other fields. The distribution of journals can, to some extent, reflect the trend of NER research and the subject areas involved. According to the titles of publications, NER-related studies have been published in 507 journals, and the names of the eight journals with high article loads and their subject areas are summarized in Table 1. They are mainly published in IEEE ACCESS, JOURNAL OF BIOMEDICAL INFORMATICS, BMC BIOINFORMATICS, and other journals. From the perspective of the journal’s field, NER’s research is mainly related to computer science, medicine, biology, chemistry, and other disciplines, consistent with the research directions analysis results. The rapid development of NER technology in specific fields such as biology, medicine, and chemistry may be related to a large number of labeled databases, high-quality labels, wide data range, and high application value in these fields, which are conducive to the development of NER.

4. Analysis of cooperation

4.1. Analysis of author collaboration network

In order to more accurately track and analyze the annual research dynamics and development trends in the NER field and ensure the timeliness of the analysis, we selected the time-slicing unit as one year in CiteSpace, which means that the retrieved documents will be carefully divided according to each year. The node type “Author” represents the analysis of the number of documents sent by authors and the cooperation between them. In order to focus on the core authors and their cooperation models who are highly active and influential in the field of NER, we will provide a highly targeted and clear visual presentation and analysis basis. Set the node label “Threshold” to 5, which will display author labels with more than five published articles. Through this setting, the visualized knowledge map of the author’s cooperation network reveals this field’s core researchers and cooperation networks, as shown in Fig. 5.

In Fig. 5, the number of articles published by an author is represented by the node’s size, where a larger node indicates that the author has published more articles. The thickness of the line between the nodes indicates the cooperation between authors. The thicker the line, the more frequent the collaboration. On the contrary, the narrower the line, the less collaboration. The node color indicates when the author published a paper, and the warm color indicates when the author published an article recently. The node size and line

Table 1

Number of papers in each journal and its subject fields.

Publication titles	Number of publications	Discipline domain
IEEE ACCESS	95	Computer Science Information Systems, Engineering Electrical Electronic, Telecommunications
JOURNAL OF BIOMEDICAL INFORMATICS	84	Computer Science Interdisciplinary Applications, Medical Informatics
BMC BIOINFORMATICS	79	Biochemical Research Methods, Biotechnology Applied Microbiology, Mathematical Computational Biology
APPLIED SCIENCES BASEL	65	Chemistry Multidisciplinary, Engineering Multidisciplinary, Materials Science Multidisciplinary, Physics Applied
BIOINFORMATICS	48	Biochemical Research Methods, Biotechnology Applied Microbiology, Computer Science Interdisciplinary Applications, Mathematical Computational Biology
BMC MEDICAL INFORMATICS AND DECISION MAKING	46	Medical information
DATABASE THE JOURNAL OF BIOLOGICAL DATABASES AND CURATION	42	Mathematical Computational Biology

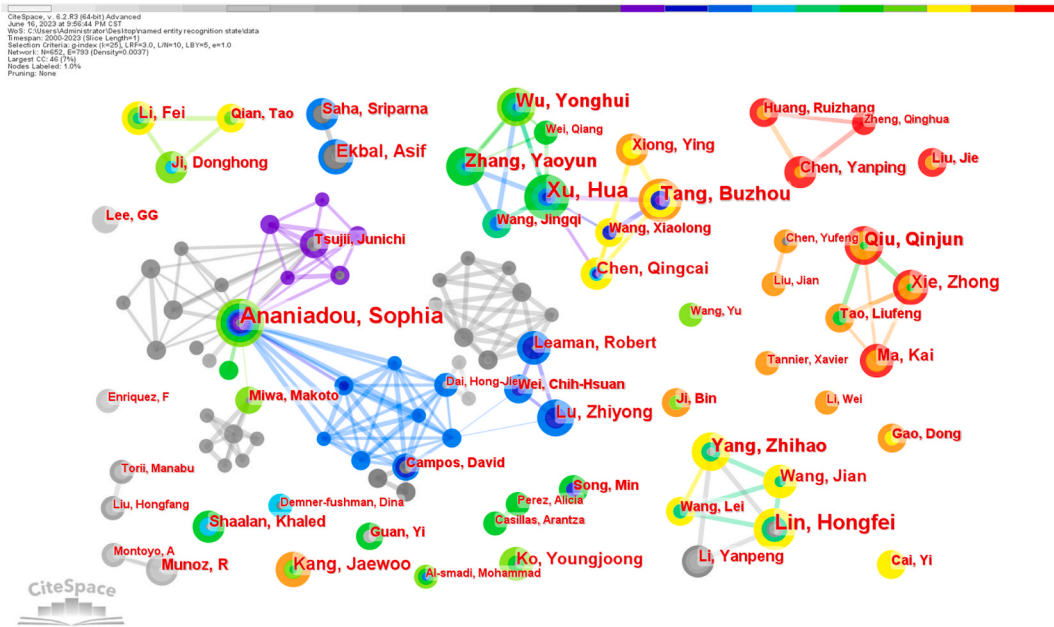


Fig. 5. Author cooperation network map.

of the author, institution, and country cooperation network map have the same meaning. In the early stage, “Munoz, R” and “Li, YP” appeared with high frequency, and the links between each node were dense, and the authors cooperated closely. In the middle period, “Ananiadou, S” and “Xu, H” appeared more frequently, and the cooperation between the authors became closer. Recently, “Lin, HF” and “Qiu, QJ” appeared more regularly, but the cooperative relationship between the authors was relatively reduced. Table 2 counts the authors with more than eight publications. Among them, “Ananiadou, S” has published the most papers and is mainly engaged in the research of Mathematical Computational Biology and Biotechnology Applied Microbiology [36,37]. The second is “Xu, H”, who published 13 papers and mainly engaged in research on Health Care Sciences and Services Medical Informatics [38,39]. It is closely followed by “Lin, HF” with 11 publications, mainly engaged in Computer Science and other research work [40,41]. The recent larger node is “Qiu, QJ” The author published the first article in 2019 and has published eight articles so far. The author is mainly engaged in research in astronomy and astrophysics, geology, and other research work [42,43], indicating that the author has recently paid close attention to the field of NER. The method for academia to determine the core authors in a field can be calculated by Price’s law, as shown in Eq. (2) [44].

$$M = 0.749 \times (N_{\max})^{1/2} \tag{2}$$

where M is the threshold for judging the number of papers published by core authors, and those whose paper count is greater than this value are core authors, and N_{\max} is the highest number of papers published by author in this field. The number of core authors was 63, and 310 articles were published, accounting for 23.85 % of the total literature numbers, which was far lower than the conclusion proposed by Price’s law that half of the papers were produced by high-productive authors, indicating that the scale of cooperation between authors was relatively small, and no core cluster was formed. Therefore, the cooperation between the authors or the author team should be strengthened.

Table 2
Author with more than (or equal to) 8 publications and the year of first publication.

Author	Year of first publication	Number of published papers
Ananiadou, Sophia	2008	15
Xu, Hua	2014	13
Lin, Hongfei	2007	11
Tang, Buzhou	2014	9
Yang, Zhihao	2008	8
Zhang, Yaoyun	2016	8
Wu, Yonghui	2016	8
Qiu, Qinqun	2019	8

4.2. Analysis of national and institutional cooperation network

There are specific differences in NER technology between languages, and the difficulties of NER for different languages are different. For example, in the NER technology of English and Chinese, the first problem the NER of Chinese faces is correctly segmenting the words in the text. English words have obvious boundary conditions, while Chinese boundary conditions are difficult to determine. Analyzing cooperation among various countries can promote the exchange and cooperation of NER technology in different languages and the development of NER technology. Yu et al. [45] explored the evolution of collaboration in the analytic hierarchy process research field through bibliometric methods, revealing the dynamic changes in collaboration between countries/regions and institutions and how these collaborations promote the sharing and diffusion of knowledge in the analytic hierarchy process field. Using this research as a reference, this article uses CiteSpace software to analyze national cooperation relationships. Optimizing the principles of map information density and clarity also ensures that the analysis can focus on countries with strong cooperative influence in the NER field, thereby effectively displaying the cooperation networks of these countries in the field. Set the node label “Threshold” to 12. The country labels with a published volume greater than 12 will be displayed. The national cooperation network knowledge graph drawn by this method (Fig. 6) counts the top ten countries with the most published articles and their betweenness centrality values (Table 3). It intuitively demonstrates the current status and characteristics of cooperation between countries in the NER field. Betweenness centrality is a measure of a node’s centrality in a network. It equals the shortest paths from all vertices to all others that pass through that node [46]. The larger the amount of data passing through the node and the more frequent the data transmission, the greater the influence of the node in the network graph and the more critical the node’s position. Fig. 6 and Table 3 can be used to understand the strength of cooperation between countries and the development status of NER technology in various countries. Among the retrieved data, PEOPLES R CHINA published the most papers, reaching 563, with a betweenness centrality value of 0.32, indicating that NER technology is developing rapidly in China and attracting a high degree of attention. The second is the USA, with 521 publications and a betweenness centrality value of 0.30. This is followed by ENGLAND, with 88 publications and a betweenness centrality value of 0.27. In addition to the countries mentioned above, the development of other countries, such as Germany, India, and Japan, cannot be ignored, although the relatively small number of publications has made important contributions to specific NER technology fields, such as multilingual recognition, cross-domain applications, etc. This demonstrates the diversity and extensive collaboration in global research on NER technologies. The number of articles published by a country reflects the development of NER technology in the language used in that country. From the number of articles published in each country, it can be seen that research on NER technology in Chinese, English, and Arabic is significantly active. At the same time, we have also noticed that NER technology in other languages, such as Spanish, French, and German, is also developing rapidly. These languages show their uniqueness in the process of word embedding, and the integration of this feature in the pre-trained language model (PLM) helps enrich the model’s understanding of the language, allowing the model to learn more features. These developments highlight the potential and wide range of applications of NER technology in adapting to global multilingual environments.

Analyzing the cooperation between institutions can help understand the leading institutions and mainstream research objects in the field. In order to focus on displaying the leading institutions with more than six publications in the field of NER and simplify the network map to highlight these major research centers and their cooperation models, the node label “Threshold” is set to 6. Through

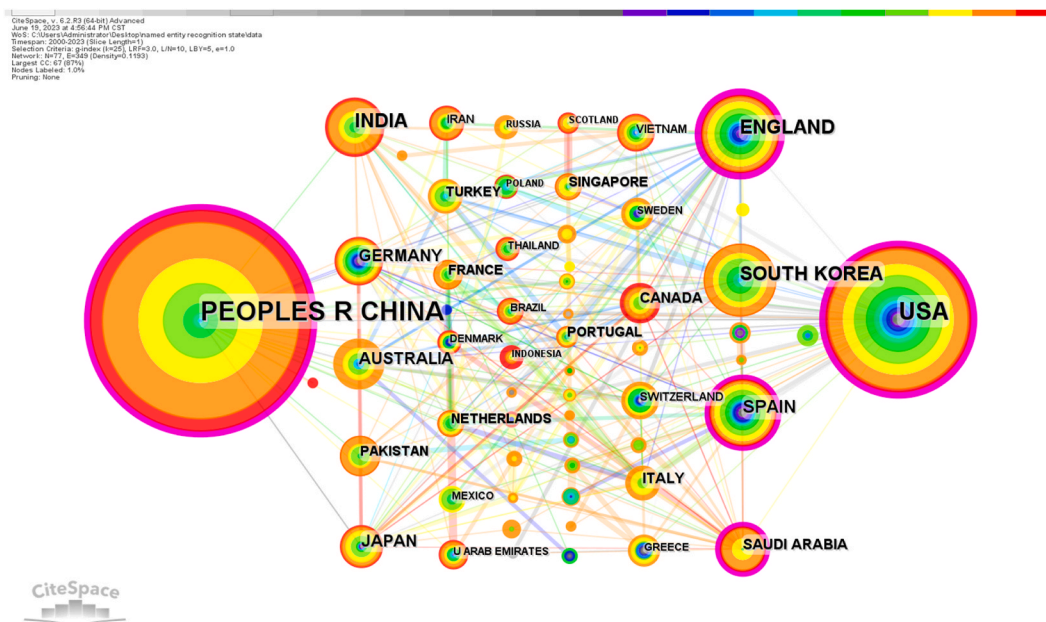


Fig. 6. map of national cooperation networks.

Table 3
Top 10 countries and institutions with published papers and their betweenness centrality.

Country	Number of publications	Betweenness centrality	Organization	Number of publications	Betweenness centrality
PEOPLES R CHINA	563	0.32	Chinese Acad Sci	34	0.19
USA	251	0.36	Harbin Inst Technol	29	0.08
ENGLAND	88	0.27	Dalian Univ Technol	25	0.01
SOUTH KOREA	85	0.02	Natl Univ Def Technol	24	0.03
INDIA	66	0.07	Univ Manchester	22	0.08
SPAIN	63	0.14	Univ Cambridge	14	0.01
GERMANY	44	0.07	Wuhan Univ	14	0.01
AUSTRALIA	37	0.06	Peking Univ	13	0.07
JAPAN	35	0.06	Korea Univ	12	0.05
ITALY	28	0.06	Cent South Univ	12	0.01

this setting, we can more clearly identify and analyze active institutions in the NER field and their cooperation networks and draw an institutional cooperation network map (as shown in Fig. 7). Table 3 lists the top ten institutions and their betweenness centrality. It can be seen from Fig. 7 and Table 3 that the institution with the most significant number of publications is Chinese Acad Sci (Chinese Academy of Sciences), with 34 publications, and its betweenness centrality value is 0.19, indicating that Chinese Acad Sci has a more significant academic influence on the field of NER; The second largest publication numbers is Harbin Inst Technol (Harbin Institute of Technology), with 29 publications, and the betweenness centrality value is 0.08; the third publication amounts is Dalian Univ Technol (Dalian University of Technology), with 25 publications, the betweenness centrality value is 0.01. It can be seen that most institutions cooperate closely. Still, most institutions are colleges and universities, so there needs to be more cooperation between schools and enterprises, and the number of papers published by enterprises is relatively small. Therefore, collaboration and exchanges between schools and enterprises should be strengthened to promote NER’s more profound development and its application at the enterprise level.

5. Literature analysis

5.1. Literature co-citation analysis

Literature co-citation refers to the co-occurrence of two or more documents in reference to one or more other documents. Literature with many co-citations in the field is vital or core literature. The analysis of literature co-citation in the area of NER can explore the mainstream models and application fields of NER technology at various stages. Table 4 lists the top 10 co-cited literature and the year of publication, which are of great significance to the development of NER technology. The CiteSpace software is used to visualize the co-citation of literature. Based on the preliminary analysis of the frequency distribution of the data used, the aim is to balance the level of detail of the atlas with the readability of the overview. At the same time, in order to accurately highlight the widely cited and

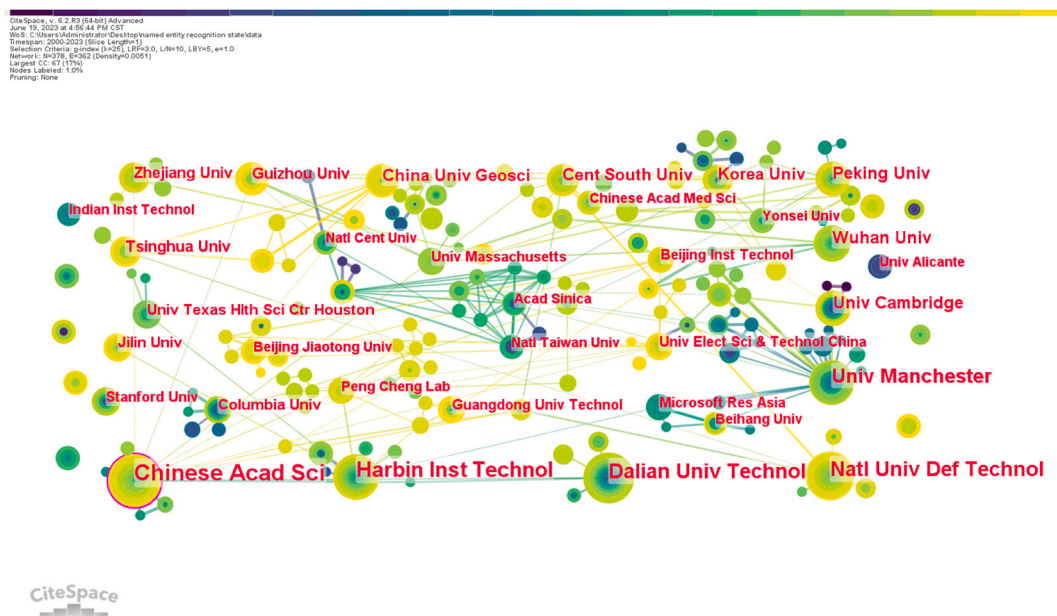


Fig. 7. Institutional cooperation network map.

Table 4
Top 10 co-cited literature and co-citation frequency.

Co-cited literature	Frequency	Year of publication
Devlin J, 2019, ARXIV, V0, P0	381	2019
Lample G., 2016, ACL, V0, PP260	195	2016
Vaswani A, 2017, ADV NEUR IN, V30, P0	169	2017
Ma XZ, 2016, ACL, VOL 1, P1064	140	2016
Lee J, 2020, BIOINFORMATICS, V36, P1234	110	2020
Habibi M, 2017, BIOINFORMATICS, V33, P137	91	2017
Zhang Y, 2018, (ACL), VOL 1, P1554	87	2018
Radford A., 2018, P 2018 CNAM CHAPT, V0, P0	84	2018
Bojanowski P., 2017, T ASSOC COMPUT LING, V5, P135	80	2017
Li J, 2022, IEEE T KNOWL DATA EN, V34, P50	69	2022

influential literature in the research field of NER. Set the node label “Threshold” to 6; document labels with reference frequency greater than six will be displayed. The document co-citation network knowledge graph drawn in this way (shown in Fig. 8) highlights the key documents that promote the progress of this field. The number of nodes in the figure is 1188. The links between nodes are 4940. The nodes’ size represents the literature’s co-citation frequency, and the larger the node, the higher the co-citation frequency. The color of the annual ring of the node is cold or warm, which means the year of publication; the cool color represents the year of publication earlier, and the warm color represents the year of publication later. The line between the nodes shows the closeness of the relationship between the two documents.

Fig. 8 and Table 4 show that. The largest node is “Devlin J. (2019)”, with 381 citations. It shows pre-training models’ profound impact and breakthrough progress in NLP research and practical applications, especially BERT (Devlin et al., 2019) (Bidirectional Encoder Representations from Transformers). This trend marks a major shift from traditional rule-based and statistical methods to deep learning-based and large-scale pre-trained models, opening a new chapter in the field of NLP. The BERT model is based on the bidirectional coding structure of the Transformer [47], and task 1 is to randomly mask some words in the input text and then predict these masked words so that the model can learn the meaning of words in the context. Task 2, the “next sentence prediction” task, predicts whether the input two paragraphs of text are consecutive texts so that the model can understand the relationship between sentences. The second node is “Lample G. (2016)”, with 19 citations. This paper [48] proposes two NER models: a bidirectional LSTM combined with CRF to capture text’s long-term dependencies and a transformation method that uses supervised and unsupervised word representation. The paper extensively uses character-level information in the NER task for the first time. This innovation provides new ideas for later processing of complex morphological languages (such as compound words in English). The third node is “Vaswani A. (2017)”. This document [47] proposes the Transformer model, and its innovative attention mechanism (Self-Attention and Multi-Head Attention) marks an important turning point in the field of NLP. Compared with traditional convolutional neural networks (CNN) and recurrent neural networks (RNN), the Transformer model greatly improves the processing efficiency through parallel

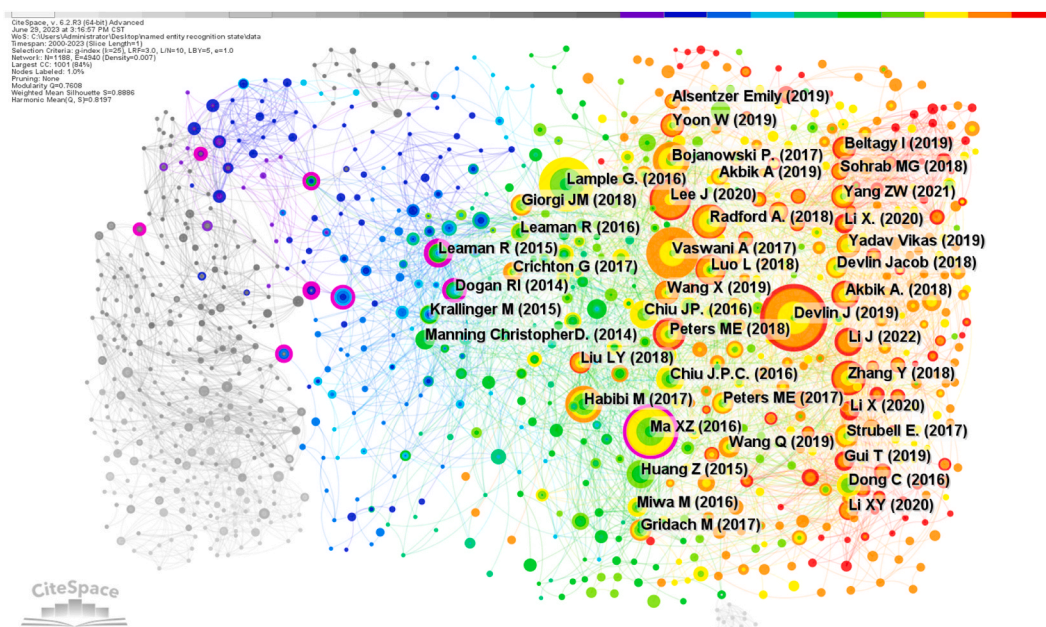


Fig. 8. Co-cited literature network map.

processing and, at the same time, captures long-distance dependencies more effectively by focusing on different parts of the text, which is the cornerstone of promoting the development of models such as BERT, and provides new methods and technical paths for solving complex NLP tasks. “Lee J (2020)” is the node with the highest citation frequency recently, with ten citations. This document [49] introduces the BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining) model, which is a BERT model optimized specifically for biomedical text mining tasks. BioBERT significantly improves the performance of tasks such as biomedical NER and relation extraction, thanks to its pre-training on a large amount of biomedical domain text. This work demonstrates how to improve the application effect of the BERT model further in professional fields through pre-training in specific fields. It opens up a new path for NLP in specific fields. The thickness of the purple growth rings of a node represents its betweenness centrality. The thicker the growth rings, the higher the betweenness centrality of the node. The node with the highest betweenness centrality is “Leaman R (2015)”. In the literature [50], the method of model combination is used to combine two independent machine learning models to create a chemically named entity recognizer, in which the two models have large differences, such as different feature sets and CRF parameters. The model’s innovative nature has led to significant advances in NER tasks in the field of chemistry.

In addition to the high-frequency co-cited literature mentioned above, other sources also play an equally important role in developing NER. For example, Sang and De Meulder [51] provide a standardized evaluation framework and dataset for NER research, which has profoundly impacted standardizing NER assessment and advancing research in this field. Collobert et al. [52] propose an end-to-end deep learning approach for handling NLP multi-tasks, such as NER and part-of-speech annotation. This approach uses a unified neural network model, reduces the reliance on task-specific feature engineering, and lays the foundation for a multi-task NLP solution. Dai and Le [53] propose a learning strategy that combines unsupervised pre-training and supervised fine-tuning, which can effectively use large amounts of unlabeled data to improve the model’s performance on specific tasks. This approach inspired later PLMs such as BERT. In addition, many other studies have made key contributions to the development of NER, including exploring different algorithms and models, applying NER in various languages and domains, and innovative approaches when dealing with complex entity types. Together, these studies have advanced NER technology, making it a key component of the field of NLP.

5.2. Cluster analysis of co-cited literature

Further clustering analysis of co-cited literature can explore the research frontier in the field of NER. The Log-Likelihood Ratio algorithm is selected as the clustering algorithm, which can effectively reflect the relationship between events. The main idea of the literature can be roughly summarized through the abstract, so the cluster label selection is obtained through the abstract. After clustering, 16 clusters are obtained, and the most significant 10 clusters are selected for display. Finally, the clustering network map of co-cited literature in the NER field is accepted, as shown in Fig. 9, modularity $Q = 0.76$ and silhouette $S = 0.89$. Q value was more significant than 0.3, meaning that the network clustering is obvious. The greater the value, the better the cluster obtained by the network. It is generally considered that $S > 0.5$ clustering is reasonable and $S > 0.7$ clustering is convincing. It can be seen that the matching relationship between the node and the cluster is high, the matching relationship with other clusters is low, and the clustering effect is good. Table 5 lists the size, average year, and cluster labels and their log-likelihood values for each cluster, with larger values indicating more representative labels. The largest cluster was the “#0 single-task” model, the cluster size was 172, the silhouette value

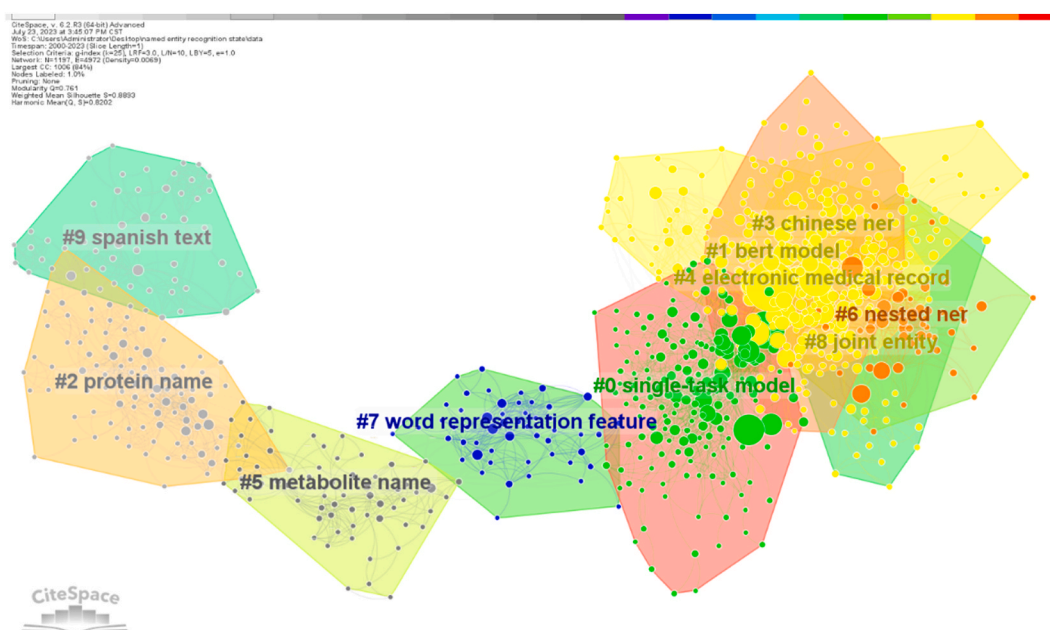


Fig. 9. Co-cited literature clustering network map.

Table 5
Co-cited literature clustering labels and their size.

Cluster ID	Size	Silhouette	Mean(year)	Top Terms (log-likelihood ratio, p-level)
0	172	0.885	2014	single-task model (1793.76, 1.0E-4)
1	109	0.783	2018	bert model (1420.22, 1.0E-4)
2	100	0.925	2003	protein name (459.62, 1.0E-4)
3	96	0.847	2018	Chinese ner (1989.18, 1.0E-4)
4	92	0.817	2017	electronic medical record (1053.37, 1.0E-4)
5	63	0.894	2007	metabolite name (204.68, 1.0E-4)
6	60	0.907	2019	nested ner (1707.52, 1.0E-4)
7	52	0.95	2011	word representation feature (677.39, 1.0E-4)
8	52	0.869	2018	joint entity (1959.86, 1.0E-4)
9	52	0.993	2001	Spanish text (210.45, 1.0E-4)

was $S = 0.883$, and the average year of the literature was 2014. This is followed by the “#1 Bert model”; the size is 110, the silhouette value is $S = 0.799$, and the average year of the literature is 2018. The third cluster is the “#1 protein name”; the size of this cluster is 101, the silhouette value is $S = 0.926$, and the average year of the literature is 2018. The emergence of the “#2protein name” cluster highlights the importance and value of NER technology in specific fields, especially in the biomedical field. Identifying protein names is of great importance in biomedical research because they are key to understanding biological processes, disease mechanisms, and drug effects. In biomedical literature, clinical reports, or research papers, proteins are a common entity type, and their accurate identification and classification are crucial for information retrieval, data mining, disease diagnosis, and biological research. With the development of NLP technology, the application of NER has been extended to various fields. In addition to biomedicine, the identification of compounds and elements in the field of chemistry, the identification of institutional and product names in the financial field, and the identification of component and process terminology in the manufacturing industry have all become important directions of NER research. The development of NER in these fields not only promotes the process of informatization and intelligence in related fields but also provides technical support for extracting and managing professional knowledge.

The “#0 single-task model”, “#6 nested ner”, “#7 word representation feature”, and “#8 joint entity” in the cluster marks the key progress and research frontier of NER technology. “#1 bert model” highlights the widespread use of PLM models such as BERT in NER tasks, demonstrating the core application of advanced models in improving text recognition and processing. “#3 Chinese ner” and “#9 Arabic ner” demonstrate the special challenges and advances that represent NER technology in different language structures. In addition, the “#2 protein name”, the “#4 electronic medical record”, and the “#5 metabolite name” reveal the unique applications and development trends of NER technology in various professional fields.

The largest cluster group, “#0 Single-task model”, means that in the development process of NER, the research of single-task models occupies an important position. This type of model focuses on performing a specific task, such as identifying a specific type of entity in text. This focus allows the model to learn more deeply and adapt to the characteristics of the specific task, thereby improving performance on the task. Compared with multi-task models, the structure of single-task models is usually simpler and easier to design and implement. This simplified design helps researchers focus on improving the model’s performance on specific tasks without worrying about the complex interaction trade-offs between multiple tasks. The concise structure and focus on a single task of the single-task model enhance the interpretability of the model’s decision-making process during the NER task. This clear and understandable feature not only makes the model an important focus in theoretical research but also promotes its in-depth analysis and understanding in empirical research. However, the main limitations of single-task models include that they are often unable to handle or generalize effectively to other types of tasks different from the training task. At the same time, because they focus on one specific task during design and training, these models cannot fully exploit the potential connections or common features between different tasks. While useful for in-depth learning of specific tasks, this focus ignores the importance of connections between different tasks when understanding text. For example, in the NER task, understanding context information or syntactic structure may be helpful to the NER task. With the advancement of technology, although multi-task and complex models gradually dominate, single-task models still have an important significance in the history and development of the NER field. The top 10 literature with the highest citation frequency in the cluster are shown in Table 6. In addition to Lample G. (2016) and Ma XZ (2016), Chiu and Nichols [54] proposed a NER model that

Table 6
Top 10 co-cited documents in the largest cluster.

Frequency	Centrality	Label	Author	Year	Source
195	0.05	Lample G. (2016)	Lample G.	2016	ACL
140	0.09	Ma XZ (2016)	Ma XZ	2016	ACL (54 TH)
64	0	Peters ME (2018)	Peters ME	2018	MNLP 2018
61	0.01	Chiu J.P.C. (2016)	Chiu J.P.C.	2016	T ASSOC COMPUT LING
60	0.03	Abadi M. (2015)	Abadi M.	2015	ARXIV160304467
43	0.28	Leaman R (2015)	Leaman R	2015	J CHEMINFORMATICS
34	0.02	Krallinger M (2015)	Krallinger M	2015	J CHEMINFORMATICS
33	0.01	Leaman R (2016)	Leaman R	2016	BIOINFORMATICS
31	0.01	Manning C.D. (2014)	Manning C.D.	2014	P C EMP METH NAT LAN
30	0.04	Crichton G (2017)	Crichton G	2017	BMC BIOINFORMATICS

combines bidirectional LSTM and CNN. This hybrid architecture effectively integrates the long-term dependency capture capability of BiLSTM and CNN's character-level feature extraction capability, making the model more effective in handling morphological changes and spelling errors, which is crucial to the NER task. This work has promoted the application of deep learning technology in the field of NER, provided a new direction and benchmark for subsequent research, and proved the effectiveness of using deep learning technology to process complex NLP tasks. E. Peters et al. [55] proposed ELMo (Embeddings from Language Models), a deep PLM based on BiLSTM. The key innovation of ELMo lies in introducing deep upper and lower cultural word embedding, which can generate dynamic word representation for the same word in different contexts. This innovation has greatly improved the performance of a variety of NLP tasks, including NER, emotion analysis, and question-answering systems. The emergence of ELMo has profoundly impacted subsequent NLP research, paving the way for developing more advanced PLMs such as BERT and GPT. The research frontier is the seed of scientific and technological innovation, which is of great significance to scientific research and economic development [56]. Through careful analysis of the recent cluster and literature co-citation and literature, we summarized the research frontier of NER technology and tried to reveal the latest scientific exploration and technological breakthrough in this field.

5.2.1. Pre-trained language model (PLM)

PLM plays a key role in the development of the NLP field. These models learn language's basic structure and patterns through pre-training on large amounts of text data, thereby understanding natural language. BERT, as one of the representatives of the PLM, obtained the note through literature clustering, which indicates that the PLM is a research hotspot of NER at this stage. BERT is the first deep, two-way, unsupervised language representation, which uses only a large text corpus for pre-training and combines the context of each token [57]. Further, fine-tune BERT through an additional output layer to apply to various downstream tasks, including NER. In addition to the flexibility of fine-tuning, BERT also has an outstanding ability to deal with rare or new words, which is particularly important in the NER task in specific fields (such as medical treatment, law, etc.). In addition, the multilingual version of BERT supports processing multilingual or cross-language NER tasks. Although BERT performs well in many aspects, it still has challenges processing long texts, such as large computing resource requirements, poor model interpretability, and great demand for fine-tuning data. These challenges have inspired various BERT improvements. For example, the RoBERTa proposed by Liu et al. [58] is an improved version of the BERT model. Optimize performance by increasing the amount of training data, using larger batches, and extending training time. The model eliminates the next sentence prediction task and introduces dynamic mask technology, allowing it to handle longer text sequences and improve understanding of complex structures. Although RoBERTa has improved in performance, its large model size may lead to increased difficulty in deployment, especially in environments with limited resources, and may face overfitting problems on small-scale datasets. Lan et al. [59] proposed ALBERT, an optimized BERT model, in order to solve the problems of memory limitation and longer training time when increasing the model. ALBERT uses two techniques: factored embedding parameterization, which reduces the model's size by decomposing the vocabulary embedding matrix. And cross-layer parameter sharing to reduce the number of parameters that increase with network depth. Like RoBERTa, ALBERT also removes the Next Sentence Prediction task. These innovations enable ALBERT to reduce the model's size and training duration significantly while maintaining a performance similar to that of BERT. Therefore, ALBERT is suitable for the fields with limited resources and provides valuable ideas for optimizing large-scale PLMs. However, although the model size is reduced through the parameter-sharing mechanism, this may also limit the model's ability to capture complex features. Furthermore, like other models, the interpretability of ALBERT remains a challenge. Recently, many researchers have fine-tuned or added other structures to perform NER tasks based on BERT and its improved series of models. Agrawal et al. [60] adopted a transfer learning method to deal with the challenge of nested named entity recognition (NER). Through joint label modeling technology, strategies such as fine-tuning, pre-training, and BERT-based language models were applied to solve this problem. Chen et al. [61] proposed a method based on the ALBERT model to extract entities from steel e-commerce data. Li et al. [62] pre-trained BERT on an unlabeled Chinese clinical record corpus and obtained a large pre-trained BERT model for Chinese clinical texts.

In addition to the BERT model, there are a series of advanced PLMs. For example, GPT (Generic Pre-trained Transformer) series models developed by OpenAI are unsupervised language representations based on deep self-attention mechanism, and the infrastructure is also a transformer. Unlike the BERT model, which focuses on improving the accuracy and depth of language understanding, GPT [63] performs better in generating tasks. At the same time, GPT uses a one-way (forward) attention mechanism compared to BERT's two-way attention mechanism, which makes the model architecture relatively simple. In addition, because of its generative nature, GPT is more flexible in dealing with open problems or generative tasks. Although GPT performs well in many aspects, it is still challenging in terms of high computational resource requirements when processing long texts, poor model interpretability, and large fine-tuning data requirements. For example, GPT-3 [64] can process longer text sequences, which helps the model understand more complex structure texts. However, its parameter scale is huge (175 billion parameters), which may lead to difficulties in deployment. In order to solve these challenges, researchers have been exploring ways to improve the efficiency of the GPT model, for example, by optimizing the model architecture, reducing the number of parameters, adopting more efficient training techniques, etc. XLNet [65] is an advanced PLM jointly developed by CMU and Google Brain. It is the first in-depth bidirectional language representation model that combines autoregressive and autocoding technologies. XLNet uses only large text corpora for pre-training and combines the context information of each tag at the same time. Like GPT, although XLNet performs well in all aspects, it is still a challenge in terms of computing resource requirements, handling long text, and fine-tuning. In addition to the models introduced in the appeal, there are many advanced models such as Transformer-XL [66], ERNIE [67], ELECTRA [68], etc. At this stage, the PLM, while achieving significant NLP capabilities, is also faced with the challenges of high demand for computing resources and environmental impact, as well as the problems of unfair and inaccurate model output that the bias of training data may cause. In addition, the lack of interpretability and limited generalization ability of these models are also the main problems. Many researchers

are also using PLM, such as GPT [69,70] and XLNet [71,72], to achieve advanced performance on NER tasks.

5.2.2. Cross-language NER and cross-domain NER

The development of NER technology in various countries and document clustering labels show that NER technology has become a global research hotspot, and its application and research scope span multiple fields and multiple languages. The development of cross-language NER makes it possible to realize efficient NER by drawing on the data and models of high-resource languages when facing some low-resource languages that lack a large amount of training data. It can reduce the need to develop and train models for each language separately, saving time and resources. At the same time, it can also promote cultural exchanges, assist in analyzing and processing multi-language documents, and provide support for information extraction and data analysis on a global scale. However, cross-language NER research also faces a series of challenges. For example, there are differences in grammar, vocabulary, and cultural background between different languages, and different models perform differently in each language. It is a challenge to find a universal model suitable for multiple languages. On the other hand, model recognition of entities often depends on context, and texts in different languages may have different contextual structures and cultural meanings, which greatly increases the difficulty of the model's understanding of text when there is only a small amount of data for fine-tuning. In addition, the consistency and quality of annotations also affect the models' training and evaluation. Datasets in different languages may differ in entity definitions and annotation standards. For the above difficulties, the researchers proposed some solutions. For example, Google launched Multilingual BERT (M-BERT), which aims to handle the task of natural language understanding in multiple languages rather than a single language by pre-training a model with the same architecture as the BERT model on a large number of multilingual texts. Facebook AI proposed XLM-R (Cross linguistic Language Model RoBERTa) [73] for cross-language NLP tasks based on the RoBERTa model architecture. The model is pre-trained on texts in more than 100 languages and uses a self-supervised learning method that does not rely on the parallel corpus, so it can better handle the differences between languages and improve model performance in multiple languages. In addition to improving the PLM, other researchers have also realized the cross-language NER task through different methods. For example, Keung et al. [74] added language adversarial tasks when fine-tuning multilingual BERT, successfully improving the model's performance in zero resource cross-language environments. However, the paper did not delve into the potential limitations of adversarial training, such as adaptability in different language combinations or more complex language scenarios. Feng et al. [75] propose three innovative strategies to improve NER task performance on low-resource datasets: transferring knowledge from high-resource languages, expanding dictionary strategies, and integrating cross-language universal word level entity type features into neural network architectures. Although many scholars have explored cross-lingual NER tasks, some difficulties still have not been resolved. For example, existing models have insufficient generalization capabilities when dealing with new languages that are significantly different from the training data. In addition, the differences in grammar structure, vocabulary usage, and cultural background between different languages still challenge the model's adaptability. Future research may focus on developing general models that are more adaptable to different languages and cultures to address these challenges. This includes exploring effective ways to transfer knowledge from resource-rich languages to languages with fewer resources and using unsupervised learning techniques to solve the problem of insufficient annotation data, thereby improving the performance of NLP-related tasks. The cross-lingual NER task proposes solutions for entity recognition in specific languages or cultural backgrounds and has profound implications for developing NLP.

Cross-domain NER involves identifying and classifying entities in multiple fields (such as healthcare, law, finance, etc.) [76]. Unlike traditional NER, cross-domain NER aims to develop a universal model that can adapt to text characteristics and entity categories in different fields. Research on cross-domain NER technology can more accurately extract key information from texts in different fields and provide support for various complex NLP applications. The difficulties in implementing cross-domain NER tasks include the model's ability to understand domain-specific knowledge in different fields, where the text may contain unique entity types and specialized terminology. Moreover, there may be significant differences in text style and structure in different fields, which poses a challenge to the model's generalization ability. Furthermore, some domains may lack sufficient annotated data to train effective NER models. Jia et al. [77] combine the transfer learning method and use cross-domain language models as a bridge to perform cross-domain and cross-task knowledge transfer, thereby solving problems such as resource limitations and domain adaptability in cross-domain NER tasks. Chen et al. [78] alleviate the problem of data scarcity in cross-domain NER tasks by using data augmentation methods such as pseudo-annotated data and data synthesis. Brack et al. [79] process data from different scientific fields simultaneously through multi-task learning methods, thereby improving the model's generalization ability. In addition to the above methods, other researchers [80,81] have also addressed the challenges faced by cross-domain NER tasks using different methods. Although there have been many studies and solutions for cross-domain NER tasks, there are still some challenges. For example, adaptability to highly specialized fields, changes within the field (new entity type terms may appear in some fields over time), small sample learning, etc. Future research may need to explore more efficient few-shot learning methods, develop more flexible model architectures, and improve domain adaptation techniques to improve the problem.

5.2.3. Nested NER and fine-grained NER

Nested named entities refer to entities that can contain or be embedded within another entity. For example, in the entity "Peking University", "Peking University" itself is an organizational entity, and the "Beijing" contained within it is a geographical location entity; this indicates that the same text fragment can be classified into multiple entity types. The traditional flat NER method cannot recognize overlapping or nested entities, but it often contains complex entity structures in medical literature, legal documents, scientific papers, and other texts. Katiyar and Cardie [82] pointed out that nested NE is quite common: 17 % of entities in the GENIA corpus are embedded in another entity; In the ACE corpus, 30 % of sentences contain nested entities. The development of nested NER technology can enable deeper analysis and understanding of text. The nested NER model has better understanding and processing

capabilities when processing highly structured or specialized text. The complexity of nested entities makes nested NER more challenging and practical than traditional NER, so nested NER has become an emerging topic in NER tasks [83]. In addition, determining the exact boundaries of each entity in nested entities is also a challenge, especially when the entities overlap and the context is ambiguous. Furthermore, when dealing with nested entities, the structure of the model is often more complex, which may result in higher computational complexity and lower processing efficiency. These challenges require continuous research and innovation in model design, data processing, and algorithm optimization to improve the performance and applicability of nested NER technology. From the perspective of model structure, standard mainstream methods for nested NER include early rule-based methods [84], which rely on the post-processing of rules like traditional NER methods. The Layer-based approaches [85] treat nested NER tasks as multiple traditional NER tasks and identify nested entities layer by layer. The Span-Based approach [86] solves the problem of entity boundary ambiguity by calculating the span representation of all sequences and then classifying them through local normalization scores. The hypergraph-Based approach [87] refers to using hypergraphs to represent the nested structure of entities in sentences and can represent and process complex entity relationships. The Transition-Based approach [88], inspired by the transformation-based parser, processes nested entities through sequential operations and is suitable for long sentences with complex structures. Recently, many scholars have addressed the nested NER problem by further improving mainstream methods. Geng et al. [89] proposed a novel planar sentence representation and bidirectional two-dimensional recursive operation, effectively solving the semantic dependency and entity boundary ambiguity problems in nested NER. This method can reduce the complexity of the model and improve the accuracy of entity recognition, but there may be a dependency on high-quality annotated data. Cui and Joe [90] proposed a pyramid hierarchical model based on a multi-head adjacent attention mechanism, which is used to fuse information from two adjacent inputs and better model the dependency relationship between entity spans. Chen et al. [91] improved the accuracy of entity boundary recognition and semantic dependency construction in nested NER by proposing a controlled attention mechanism, allowing the model to focus more effectively on task-related semantic features, thereby improving the model's performance and robustness. Although many researchers have addressed some of the difficulties in nested NER through various methods, some challenges still need to be addressed. For example, existing models still underperform when dealing with extremely complex nested structures, such as multiple nested or cross-nested entities. Secondly, many models perform well in specific fields, but their performance may decrease when applied to different types of text or across domains. Meanwhile, in low-resource languages, how to effectively identify nested NER is also a challenge. Therefore, future research on nested NER technology may focus on the following points. We will use weakly supervised and transfer learning techniques to reduce dependence on large amounts of annotated data and improve the model's adaptability in different fields. Explore nested NER methods that combine multimodal data such as text, sound, audio, and cross-language nested NER. Develop more efficient and lightweight nested NER models to meet real-time and large-scale data processing needs.

Fine-grained NER aims to identify and classify entities using more detailed and specific categories from text. Fine-grained NER focuses more on more profound and specific entity categories than traditional NER. For example, it identifies an entity as an organization and further distinguishes it from government agencies, educational institutions, commercial companies, etc. This requires the model line to understand the context more deeply to accurately classify close or similar entity types. At the same time, it will also face problems such as increasing entity categories, blurring entity boundaries, and fine-grained feature recognition. These challenges require that the fine-grained NER model not only needs strong language understanding ability but also can handle complex entity relationships and category segmentation. Rodríguez et al. [92] effectively address challenges such as blurred entity boundaries, inaccurate category recognition, and complex context interpretation in fine-grained NER by combining advanced text encoding technology, BiLSTM, CRF, and name-focused attention mechanisms. Wan et al. [93] propose a span-based multimodal attention network, which introduces a closed-loop mechanism to simulate human behavior to simultaneously and deeply mine multimodal information (span cell tag sequence and context information) existing in the text to capture the fine-grained interaction characteristics between them, thus improving the model performance. Wang et al. [94] use the method of distance-supervision, combined with flexible knowledge base matching and ontology-guided multi-type disambiguation technology, to effectively deal with the fine-grained NER problem in the chemical field. The performance of fine-grained NER can be improved through advanced disambiguation technology, combining NER with entity linking to enhance the understanding and classification of complex entities, and also by utilizing data synthesis, transfer learning, and other technologies to address issues related to data scarcity and imbalance.

5.2.4. Multimodal NER

Multimodal NER is a technology that combines text with information from other modalities (such as images, videos, sounds, etc.) for entity recognition. It analyzes the text's language features and uses information from other modalities to assist in the recognition and understanding of entities. For example, a multimodal NER system might combine visual cues in images and image description text to identify specific tasks or objects in images. When the context information is ambiguous, multimodal NER can more accurately identify entities that are difficult to determine in the text by combining multimodal data. In addition, when dealing with text containing complex scenes (such as social media content), multimodal information helps to better interpret and understand entities. When carrying out multimodal NER, the first problem is the high cost of multimodal data acquisition and annotation. Secondly, different modal data (such as text and image) may have great differences in feature representation, scale, and type, and how to effectively fuse these heterogeneous data. In addition, there may be noise inconsistency between different modes, which may affect the model performance. Yu et al. [95] realize the effective fusion of text and visual information by combining a unified multi-mode converter and an auxiliary entity range detection module. It has improved the problem of dealing with visual bias and modal interaction, thus improving the entity recognition rate in social media posts. Zhang et al. [96] proposed a multimodal graph fusion method to improve the effect of entity recognition in social media posts. By creating a graph structure that fuses text and visual objects, this method realizes deep semantic interaction inside and outside the mode and effectively integrates context and cross-modal content. The span-based

multimodal variational autoencoder proposed by Zhou et al. [97] solves the difficulty of obtaining and labeling data sets and the noise problem. The reliance on large amounts of labeled data is reduced through semi-supervised learning, and the noise in the data is effectively handled through variational autoencoders. On this basis, future multimodal NER research will further explore cross-domain and cross-language adaptability to improve the generalization ability of models in different environments. At the same time, it will also focus on developing multimodal NER systems that adapt to real-time and dynamic environments, such as social media analysis and real-time news processing, to meet the growing demand for real-time data processing.

5.2.5. Few-shot NER

In order to solve the problem of limited annotation data and high annotation cost in specific fields or low resource languages and improve the generalization ability of models when facing new entity types and different fields, the few-shot NER technology came into being. Developing few-shot NER technology can reduce the model's dependency on a large amount of labeled data, thus lowering the cost associated with collecting and labeling vast amounts of data. In addition, few-shot NER allows the model to quickly adapt to new domains, which is particularly important in dynamic environments. Driven by a small amount of data, new methods, model architectures, and algorithms such as meta-learning and transfer learning have been developed. How to enable the model to learn from a small number of samples and effectively generalize to unseen entities, solve the negative impact of noise (such as error labeling) in a small number of samples on model performance, and deal with different language styles and entity types of domain texts are the primary challenges in few-shot NER. Wang et al. [98] introduced a data enhancement method to improve few-shot NER, which enhances model generalization and training effects by changing the prompt order. Chen et al. [99] propose a self-describing network that learns extensive knowledge through pre-training and then transfers to few-shot NER tasks. This approach uses universal concept descriptions to automatically map new entity types and identify entities adaptively. Das et al. [100] propose a comparative learning method to optimize the distribution differences between labeled entity representations. Gaussian embedding is used to display the distribution of modeling entities. In this way, the model can more effectively capture the label dependency, avoid the overfitting problem of the previous methods in dealing with O (non-entity) markers, and thus improve the model's performance in the small sample NER. Chen et al. [101] propose a prompt-based metric learning framework, which effectively solves the problem of tag scarcity and overfitting by combining tag awareness prompts and metric learning. In addition to the above methods, some researchers [102, 103] explored different methods to improve the performance of NER tasks with few samples. In the future, the development of few-shot learning may need to study more complex and effective methods to encode and use context information to improve the recognition ability of models for complex entities. At the same time, we can use better data enhancement techniques and semi-supervised learning methods to expand the training data set according to the data situation. And develop lightweight models and computational efficiency optimization methods, especially in resource-constrained environments.

5.3. Analysis of burst literature

Important literature in the development process of the research field can be discovered through the burst function in CiteSpace software, which can be used to find the literature with the strongest citation burst. The literature has time characteristics, and its burst and blanking times can be known through this software to obtain the hotspot evolution and development track in this field. Burst literature refers to literature cited in a large number within a certain period [104]. The top 25 documents with the strongest citation

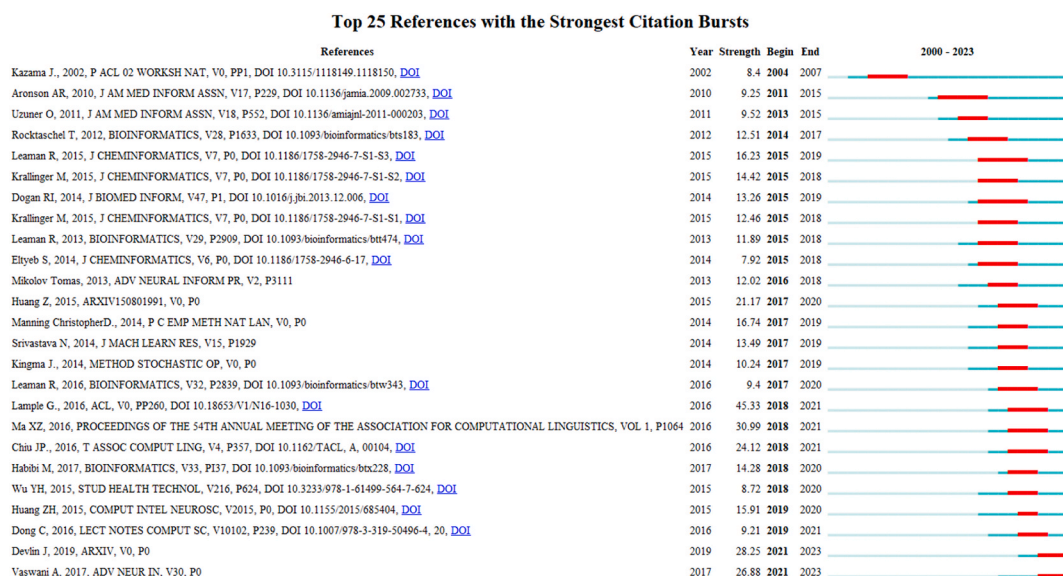


Fig. 10. top 25 literature with the strongest citation burst.

burst in the NER field are shown in Fig. 10. The blue line represents the time axis, the red line represents the period of the burst literature, and both ends of the red line represent the start and end time of each burst literature. The development of each period can be seen in this figure. For example, Kazama et al. [105] explored the application of SVM in biomedical NER. During this period, the methods of NER technology were mainly based on machine learning methods. Rocktaschel et al. [106] proposed an integration method that combines dictionary-based and grammar-based methods, effectively improving the accuracy and efficiency of extracting chemical entities from chemical texts. This method is of certain importance for specific application scenarios. Mikolov et al. [107] proposed two models for calculating word representation: the Continuous Bag of Words Common Bag of Words (CBOW) and Skip-gram models. This significantly simplifies and improves the computational efficiency of word vector representation. This document is one of the important documents that pushed the combination of deep learning and NLP tasks to the mainstream. Huang et al. [108] applied BiLSTM and CRF models to sequence marking tasks for the first time, such as part-of-speech tagging (POS), blocking, and NER, significantly improving task performance and reducing dependence on word embedding. It lays a foundation for applying further in-depth learning in the NER field.

6. Keyword analysis

6.1. Keyword co-occurrence analysis

Keywords in literature are usually the concentration of an article, and keywords can reflect the core idea of the literature. Keyword co-occurrence refers to the number of occurrences of the same keyword in a group of documents, and the close and distant relationship between them is studied by counting the number of co-occurrences. Cluster analysis can classify the keywords with strong homogeneity into one category according to the affinity between keywords, making the cohesion between keywords in the same category stronger than that between keywords in other categories [109]. Yu et al. [110] explored the knowledge structure in the field of Preference Ranking Organization Method for Enrichment Evaluations through the analysis of co-word networks, revealing the dynamic changes of the core themes and research directions in this field. This method demonstrates the effectiveness of co-word network analysis in significantly identifying and tracking the development trend of knowledge in the subject area. Based on this, this study uses this method for reference and keyword co-occurrence analysis to explore research hotspots in the field of NER in depth. The node type is selected as "Keyword", and other parameters remain in default settings. In order to enable the atlas to focus on the keywords that frequently appear in many kinds of literature and have significant importance and representativeness, optimize the information density of the atlas and ensure that important research trends and hotspots are presented in the atlas. After preliminary data exploration, we set the threshold to 26. This setting aims to emphasize the core and widely concerned research topics in the field of NER and map the knowledge map of the keyword co-occurrence network, as shown in Fig. 11. The number of nodes is 634, and the lines between nodes are 2907. The node size represents the number of keywords co-occurrences. The larger the node, the higher the number of keyword co-occurrences. The color of the annual ring of the node represents the year when the keyword co-occurrence, the cool color represents the year earlier, the warm color represents the year later, and the line between the nodes means the closeness of the two keywords. The high and low-frequency word demarcation values are calculated by Eq. (3), proposed by Donohue [111].

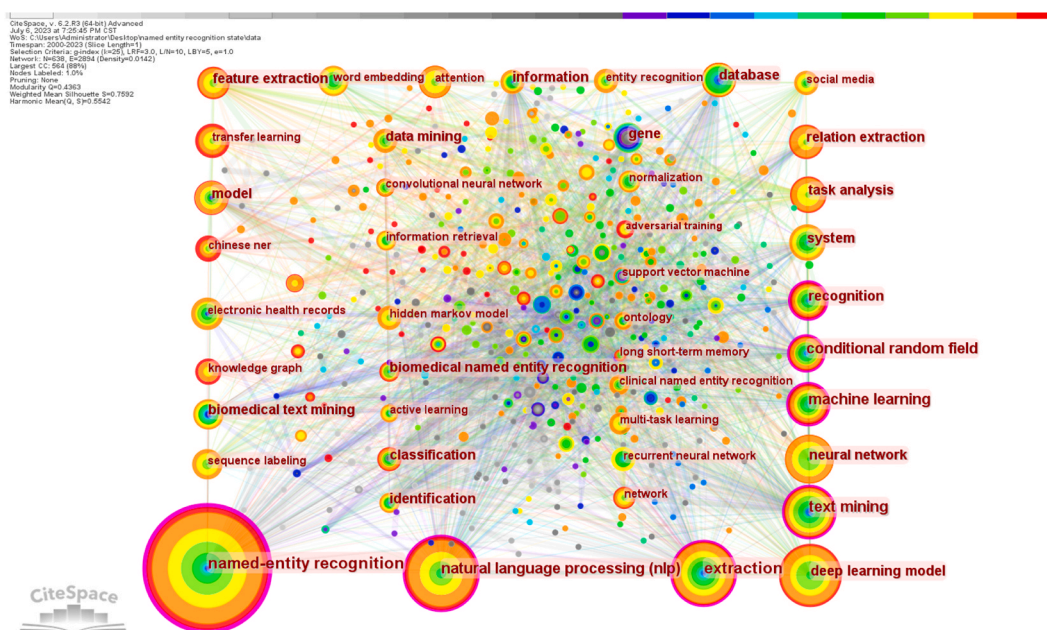


Fig. 11. Keyword co-occurrence network map.

$$T = \frac{-1 + \sqrt{1 + 8I}}{2} \quad (3)$$

where I is the number of words with a frequency of one, and T is the dividing frequency of high and low-frequency words. The number of keywords with a frequency of 1 calculated by CiteSpace is 395, and $T = 28.61$ is calculated according to the formula. Hence, those with a co-occurrence frequency greater than 29 are high-frequency keywords. The betweenness centrality, frequency, and year of occurrence of high-frequency keywords are obtained after processing the keywords with low correlation and similar relationships, as shown in Table 7.

It can be seen from Fig. 11 and Table 7 that, except for the subject words, keywords with high co-occurrence frequency in the NER field include “Natural language processing”, “Extraction”, “Text mining”, etc. These are the downstream tasks of NER. “Deep learning model”, “Neural Networks”, “Machine Learning”, etc., are methods or techniques to conduct NER research. The recent high-frequency keywords include “Task analysis”, “Attention mechanism”, “Transfer learning”, and “Feature extraction”, which show that most NER research at this stage is based on these technologies. In addition to the high-frequency keywords, attention should be paid to the keywords “Multi-task learning” and “Adversarial training”.

6.2. Keyword cluster analysis

After further clustering the keywords, 10 clusters are obtained, and the keyword clustering network map is shown in Fig. 12. Table 8 lists the size of each cluster, the mean year as well as the cluster label, and its cluster value. The largest cluster is “# adversarial learning”, the cluster size is 94, the clustering modularity is 0.711, and the average year is 2019, indicating that this strategy is widely used in recent NER technologies and is an effective means to improve the performance of NER models. The second cluster is “# social media”, the cluster size is 90, the cluster modularity is 0.695, and the average year is 2017. With the development of the Internet, more and more studies focus on NER in social media [112], which is challenging due to its informality and strong noise. The third cluster, “# biomedical literature”, has a cluster size of 90, a cluster modularity of 0.769, and an average year of 2011, indicating that NER for biological texts is a long-term research hotspot. Based on keyword co-occurrence and keyword clustering analysis, the research hotspots in the field of NER can be summarized as follows.

6.2.1. Attention network

The attention mechanism was first proposed in computer vision, which can make neural networks pay more attention to the valuable information in the input and reduce the attention to irrelevant information. Just like when people look at pictures, they tend to pay more attention to the content they are interested in. Generally speaking, the attention mechanism is divided into two steps: calculating the attention distribution on the input information and calculating the context vector according to the attention distribution [113]. The attention mechanism enables the model to focus on key parts of input data to better understand the context. In NER, it is important to understand the context around an entity because it can help distinguish between entity and non-entity terms. When dealing with long-distance dependency problems, the attention mechanism can effectively capture the dependencies between long-distance words, which is particularly important for identifying entities that span multiple words. At the same time, the attention score provides a way to explain model decisions, which can show the most concerned part of the model when identifying entities. Combining the attention mechanism with other deep learning technologies (such as LSTM, RNN, etc.) can improve the model’s overall

Table 7
High-frequency co-occurrence keywords.

Keyword	Centrality	Year of first appearance	Frequency
named-entity recognition	0.28	2004	497
natural language processing	0.18	2002	213
extraction	0.19	2002	189
text mining	0.19	2005	136
deep learning model	0.08	2017	133
neural network	0.08	2018	87
machine learning	0.11	2006	77
conditional random field	0.16	2006	68
task analysis	0.05	2019	57
database	0.06	2005	56
relation extraction	0.04	2015	55
attention	0.03	2019	39
gene	0.05	2004	36
transfer learning	0.01	2020	35
biomedical text mining	0.03	2006	35
electronic health records	0.02	2013	34
feature extraction	0.03	2019	34
biomedical named entity recognition	0.06	2004	34
classification	0.03	2007	33
word embedding	0.03	2017	32
sequence labeling	0.03	2017	30

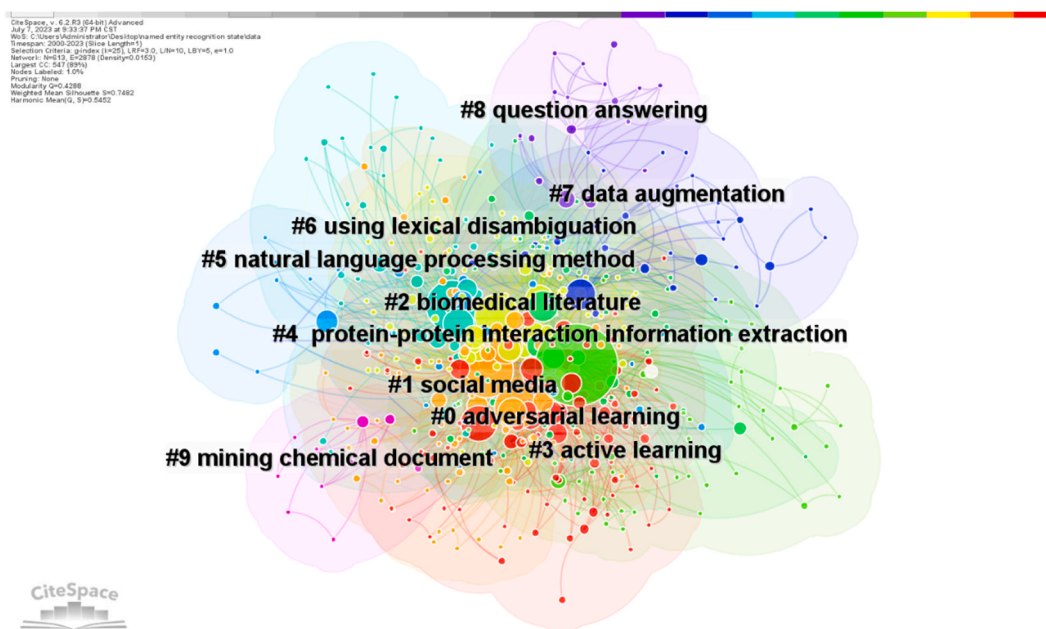


Fig. 12. Keyword clustering network map.

Table 8

Keyword clustering labels and their size.

Cluster ID	Size	Silhouette	Mean year	Top Terms (Log-Likelihood Ratio, P-Level)
0	94	0.711	2019	adversarial learning (510.39, 1.0E-4)
1	90	0.695	2017	social media (485.51, 1.0E-4)
2	90	0.769	2011	biomedical literature (1151.48, 1.0E-4)
3	60	0.644	2011	active learning (461.07, 1.0E-4)
4	59	0.749	2013	protein-protein interaction information extraction (253.17, 1.0E-4)
5	49	0.81	2013	natural language processing method (218.87, 1.0E-4)
6	35	0.812	2017	using lexical disambiguation (234.06, 1.0E-4)
7	32	0.764	2014	data augmentation (187.01, 1.0E-4)
8	19	0.896	2011	question answering (175.99, 1.0E-4)
9	10	0.958	2011	mining chemical document (60.6, 1.0E-4)

performance. In the research of attention mechanism, the main challenges include how to allocate attention accurately and efficiently in key parts of the model, especially in the face of many possible concerns to ensure the accuracy and efficiency of attention. When dealing with long sequences, how to effectively maintain the concentration and distribution of attention and avoid paying too much attention to distractions or focusing on irrelevant parts. In addition, when using the multi-head attention mechanism, optimizing the role of each head and integrating their output to improve the overall performance are key issues. These challenges highlight the complexity and nuances of applying attention mechanism to practical problems, which need in-depth research and innovative methods. Bahdanau et al. [114] used attention in NLP tasks for the first time to extract important information in sentences by giving different weights to words. With the attention mechanism demonstrating superior performance in NLP tasks, various attention mechanisms have been proposed to enhance this capability. Xu et al. [115] introduced Hard Attention, which focuses on specific parts of the input sequence to improve computing efficiency and model interpretability. However, it may face certain challenges in the training process and risk missing other important information. Vaswani et al. [47] proposed the transformer architecture, which is completely based on the self-attention mechanism and effectively handles long-distance dependencies while achieving higher parallelism. In addition, a multi-head attention mechanism is used, allowing the model to focus on different parts of the sequence on different “heads” simultaneously, thereby capturing various relationships and patterns in the data. In addition to the above attention mechanisms, many researchers improve the accuracy of NER tasks by exploring other attention mechanisms or integrating attention mechanisms into the model. Zhang et al. [116] proposed a part-of-speech attention mechanism to obtain the contribution weight of part-of-speech to entity recognition. Lin et al. [117] applied the attention mechanism to character and word level information, respectively, and proposed a neural network model that relies on hierarchical attention to achieve sequence tagging. Xu et al. [118] proposed an attention-based neural network architecture, which relieves context dependency by using document-level global information obtained from documents represented by a pre-trained bidirectional language model with neural attention. Although the attention mechanism has been widely used, the transparency and interpretability of its decision-making process still need to be

improved. In addition, using or improving the attention mechanism to deal with long texts more effectively also requires more effort. At the same time, explore the integration of different types of attention mechanisms and use their respective advantages to improve the model performance.

6.2.2. Multi-task joint learning

The core idea of multi-task joint learning is to train a model to perform multiple related tasks simultaneously rather than train a model for each task separately. The main advantage of this method is that it can enable the model to learn shared representations and features from multiple tasks, thus improving the performance of each task and the overall efficiency of the model. In NER, joint entity relationship extraction, as the representative of multi-task joint learning, allows the model to simultaneously identify entities in the text and judge the relationship between these entities, thus improving the accuracy and efficiency of information extraction and contributing to a deeper understanding and analysis of text content. Joint entity relationship extraction [119] refers to the joint process of entity recognition and relationship extraction. The joint learning method considers the potential dependency between the two tasks, thus using rich contextual information [120]. In the traditional way of extracting triple groups (entity 1, relationship, entity 2), NER and relationship extraction are executed independently, called the pipeline model [121]. The pipeline method is simple and flexible, but it ignores problems such as low-level interaction and error propagation [122]. The joint model is usually more efficient than sub-step processing because it reduces the complexity of the processing process. However, building a complex model that can simultaneously handle entity recognition and relationship extraction is necessary when extracting joint entity relationships. At the same time, data in this area is relatively scarce in many fields. Moreover, it becomes more difficult to deal with text with complex structures and diversified entities or relationship types. Zhao et al. [123] proposed a method based on a heterogeneous graph neural network, which can accurately capture the dependencies of entities and their relationships by representing iterative fusion technology and effectively dealing with entities and relationships in long texts. Chen et al. [124] used the location-aware attention mechanism and relationship embedding method to solve the problem of overlapping triples in joint entity and relationship extraction. This method improves the model's ability to deal with complex relationships by accurately identifying the location of entities and enhancing the understanding of relationships between entities. Wan et al. [93] conducted an in-depth analysis of multimodal information in the text, such as span cell label sequences and context information, and developed a multimodal attention network and a modal attention enhancement module to jointly model this information, aiming to capture the fine-grained interaction characteristics between entities and their relationships. In addition to the methods used in the above papers, some other researchers [125–127] use different strategies to improve the performance of the joint entity relationship extraction model. In addition to joint entity relationship extraction, the NER task is usually combined with the POS tagging task [128]. Combining NER with POS tagging can significantly enhance the model's understanding of words' grammatical roles and boundaries and improve the accuracy and precision of entity recognition in the context. This combination uses part of speech information to optimize entity identification and classification, especially when identifying entities in complex sentence structures. Combining NER and semantic role labeling can significantly improve the model's entity recognition ability in complex contexts. An in-depth understanding of entities' semantic roles and context relationships can enhance the understanding of sentence meaning and accurate recognition of relationships between entities. In addition to the above combinations, it is combined with syntactic dependency analysis, sentiment analysis, and text classification to improve the accuracy of each task. Although multi-task joint learning can bring significant improvement, it also faces a series of challenges. For example, task relevance, not all tasks contribute to common learning, and incorrect task combinations may lead to performance degradation. There may be conflicts between different tasks, and the optimization of one task may have a negative impact on another task. Balancing various tasks is also a key issue. It is necessary to ensure that no single task dominates the learning process. Future developments in multi-task joint learning may focus on improving the effectiveness of task combinations, such as developing advanced algorithms that can automatically adjust the learning process based on the relevance and complementarity of tasks. Optimize resource usage, such as developing dynamic resource allocation mechanisms, exploring adversarial training and regularization techniques to enhance the generalization and robustness of the model, etc.

6.2.3. Transfer learning

In most fields, tagging data is often insufficient, and tagging costs are high. Transfer learning can solve the problem of insufficient tagging data to some extent. Transfer learning [129] refers to applying the knowledge learned in the source task domain to machine learning tasks in the target domain. Transfer learning is an effective method for low-resource corpus and cross-domain learning. In the NLP field, there are generally two types of transfer learning. One is feature-based transfer, mainly represented by word2vec [107]. The other is fine-tuning. The whole pre-trained model is carried to the downstream task for fine-tuning, mainly represented by BERT. At this stage, the focus of transfer learning is mainly on fine-tuning. However, the following points need to be considered when conducting transfer learning. One core of transfer learning is effectively applying the knowledge learned from one field to another. The two fields may have significant differences in data distribution, feature space, or task objectives, making direct transfer ineffective. In addition, when the difference between the source domain and the target domain is too large, it may lead to negative transfer; that is, the knowledge of the source domain may not only be unhelpful but also harm the performance of the target domain. It is also a key point in determining which features are shared between the source and target domains and which are domain-specific. Peng et al. [130] employ a language model based on BiLSTM as part of their transfer learning approach. This model is initially trained to extract features and structures from a large corpus of unlabeled text data. These learned linguistic patterns are then adapted and applied to the task of NER. Gligic et al. [131] pre-trained the model on many unlabeled electronic health records to capture rich linguistic features and context information and then applied these embeddings to the network architecture. Through this transfer method, the information in a large amount of unlabeled data can be effectively used to improve the model's performance. Yu et al. [132] used BERT to

conduct pre-training on large-scale text corpora and the medical department's deep-level, context-based two-way language representation. Then, they used the output of BERT as the input of the subsequent neural network model. Yao et al. [133] adopted transfer and active learning methods to address the scarcity of labeled data by learning from public source datasets to knowledge transfer to fine-grained mechanical NER. Transfer learning is a rapidly developing field, and future research may continue to explore how to transfer knowledge more effectively between different fields or modalities, such as transferring from text to images or across various types of datasets. Additionally, developing technology that can automatically identify the optimal transfer strategy could reduce the need for manual adjustments and extensive experimentation.

6.2.4. Adversarial training

Adversarial training [134] is a technology used to enhance the robustness of machine learning models, especially for deep learning models. It trains the model by introducing and using adversarial samples to improve the model's resistance to small, deliberately created disturbances in the input data. Through training models to deal with diverse inputs, adversarial training can enhance the generalization ability of models to unseen data. In addition, adversarial training increases the diversity of training data by creating adversarial examples, and in some cases, not only improves the robustness of the model but also improves its performance. For security-sensitive applications, such as fraud detection, adversarial training can improve the stability of the model in the face of malicious operations. While adversarial training brings excellent performance, other factors must be considered. For example, adversarial training requires an additional computational burden in generating and training on adversarial training examples. At the same time, consideration should be given to finding a balance between enhancing robustness and maintaining high performance when conducting adversarial training. Excessive adversarial training may lead to a decline in model performance on conventional data. In addition, during the adversarial training process, if the adversarial sample generation method is too simple or is too different from the real data distribution, it may cause the model to overfit the adversarial sample. In NER, adversarial training not only improves the robustness and generalization ability of the model but also helps the model learn to recognize semantically complex or ambiguous entities, such as entities with ambiguity or context dependence. Inspired by this technology, some studies improve the performance and robustness of NER through adversarial training. For example [135], proposed a cross-domain adversarial learning method, which guides the model to learn the shared information between two tasks (Chinese electronic medical record text and NER in online medical consultation text) by combining the adversarial mechanism into multi-task learning, thus significantly improving the recognition performance and robustness of the model on complex and diverse text data. Park et al. [136] performed NER in the automotive field by combining adversarial training and multi-task learning. Adversarial training trained the model to recognize terms from both the general and automotive domains, thereby avoiding overfitting in a single domain. Multi-task learning is applied to handle NER and word spacing prediction tasks simultaneously. Wang et al. [137] increased the adaptability of the model to data by adding disturbances to the key variables of the model. The purpose of this adversarial training is to improve the generalization ability and robustness of the model, reduce the risk of overfitting, and improve the model's performance in processing diversified input data. In the field of NER, adversarial training has shown its potential and effectiveness. Future research and development may focus on using adversarial training to improve the capability of the NER system in dealing with complex entity structures (such as nested entities) and cross-domain and cross-language adaptability. In addition, it may also focus on improving the NER system's ability to deal with low-resource language and informal text, enhancing the system's security and resisting adversarial attacks.

6.2.5. Deep active learning

Active learning [138] is a method committed to studying how to obtain more performance gains as much as possible through less labeled data. Specifically, iterative unlabeled data sets select appropriate samples for labeling to reduce labeling costs. However, classical active learning methods are difficult to deal with in terms of high-dimensional data [139]. Deep learning performs excellently in processing high-dimensional data and feature extraction, and active learning can effectively reduce the annotation cost. Therefore, combining deep learning with active learning provides an effective way to train efficient models when data annotation resources are limited. In NER, the main advantage of deep active learning is that it can significantly reduce the need for high-quality annotated data, thereby reducing annotation costs and time. At the same time, the accuracy and adaptability of the model are improved by selecting the most effective samples for model improvement, especially in data scarcity or domain-specific scenarios. However, when performing active learning, selecting unlabeled samples to best improve the model's performance requires a precise sample selection strategy. At the same time, high-quality annotation may still rely on experts in specific fields, especially in professional fields. In addition, processing a large number of unlabeled data to determine its information content may lead to higher computing costs. In recent years, many scholars have achieved excellent results in the field of NER by combining deep learning and active learning. For example, Agrawal et al. [140] used the minimum confidence sampling strategy based on uncertainty to solve the sample selection problem. This strategy considers the uncertainty of the model's most likely label for each instance and calculates the number of uncertain words in the sentence. At the same time, the corpus is used to label the selected samples directly. Li et al. [141] combined the uncertainty- and diversity-based sampling method with the BERT-BiLSTM-CRF model to alleviate the problem of insufficient annotated data. Among them, uncertainty sampling selects instances with uncertain labels, while diversity sampling increases data diversity and selects instances with large context differences. Radmard et al. [142] proposed a sequence-based active learning method to improve the efficiency of sample selection in the NER task. This method not only considers the uncertainty of the whole sentence but also focuses on the subsequences in the sentence, allowing the query and annotation of subsequences with high uncertainty. In the field of NER, future deep active learning will focus more on further developing subsequence-based annotation methods to improve the utilization of annotation data and reduce the labor and time costs required for annotation. At the same time, it explores integrating the latest deep learning models (such as PLMs) and technologies into the deep active learning framework and developing more accurate sample

selection algorithms.

6.2.6. *Federal learning*

Federated learning [143] is a distributed machine learning method whose core is to allow multiple devices or servers to collaborate on data learning while protecting data privacy and security. Compared with traditional machine learning methods, federated learning can not only protect data privacy and reduce dependence on centralized data storage but also enable distributed collaborative learning across multiple devices. Driven by privacy protection and data security, the application of federated learning in NER has gradually become a research hotspot. For those NER applications involving sensitive data, such as healthcare, financial services, or legal documents, federated learning provides a way to protect personal privacy and sensitive information while allowing learning to be performed from this data. In addition, sharing models without sharing data between different institutions or fields allows the NER system to learn from a wider range of data, thereby improving the model's generalization ability and accuracy. However, when conducting federated learning, it is necessary to consider ensuring data privacy and security in the process of distributed data processing, solving the heterogeneity problem of data distribution on different clients, improving communication efficiency in the model update process, and reducing Broadband and resource consumption. Developing federated learning suitable for NER requires careful consideration of data distribution, model design, privacy protection, communication optimization, and other aspects. Wu et al. [144] use knowledge distillation technology to achieve communication-efficient federated learning, using smaller mentee models and larger mentor models to learn from each other. Small models perform personalized learning on their respective clients while reducing communication costs through the dynamic gradient method based on singular value decomposition. Wang et al. [145] proposed a cross-platform data distillation and processing method for heterogeneous label sets to train global NER models. This method combines the sequence-to-sequence NER framework and prompt tuning technology to reduce communication costs and improves through the distillation of pseudo-complete annotations (data contains all possible entity type annotations, not just the entity types already in the local dataset) Identification of unlabeled entity types. Ge et al. [146] divided the model into private modules that focus on local characteristics and shared modules that capture general knowledge of the platform, and then shared model gradients rather than original data among various medical platforms to maintain privacy, thereby improving the generalization ability and accuracy of the NER model. The research of federated learning in NER may need to explore encryption and privacy protection technologies further, such as homomorphic encryption and differential privacy. At the same time, the algorithm and processing strategy are optimized to deal with the efficiency problem in large-scale heterogeneous data environments, especially in real-time updates and dynamic changes of medical data.

6.2.7. *Distance-supervision and weakly-supervised learning*

Distance-supervision and weakly-supervised learning are two machine learning paradigms that aim to solve the problem of scarcity of annotated data. The distance-supervised learning method automatically labels training data using existing knowledge bases or external resources. In NER, the application of distance-supervised learning significantly improves efficiency and professionalism. For example, using existing knowledge bases in the medical field, medical terms in text data can be automatically annotated. This method not only reduces the need for manual annotation but also automatically identifies and accurately annotates entities in specific fields (such as medical and legal) through a professional knowledge base. In addition, distance-supervised learning can generate a large amount of diverse training data, thereby improving the model's generalization ability. There are several key points when applying distant supervised learning in NER. First, processes that rely on automated annotation in existing knowledge bases may generate erroneous or inaccurate labels, introducing noise that impacts model performance. Secondly, distance-supervised learning may have difficulty correctly handling entity ambiguities in context during automatic annotation, especially when the same words or phrases represent different entities in different contexts. In addition, since the knowledge base may cover some entity categories more extensively than others, it may lead to category imbalance in the data set, further affecting the model's ability to identify rare entity categories. Li et al. [147] proposed a self-training framework for category rebalancing. By designing flexible category thresholds and using hybrid pseudo-labeling technology, the category imbalance problem of NER under distance supervised learning is improved. Zhou et al. [148] proposed a distance-supervised learning NER method, which uses an external knowledge base to generate labels automatically and a reliability-based learning strategy to reduce false negative samples generated by incomplete labels. Meng et al. [149] automatically generate training data by matching entity mentions in the original text with entity types in the knowledge base. A noise-robust learning scheme is proposed to solve the problem of incomplete and noisy labels, including a new loss function and steps to extract noisy labels. Although distance-supervised learning has made significant progress in the field of NER, there are still some challenges. For example, for higher-level semantic understanding, distant supervised learning usually relies on surface-level text matching and simple rules, which makes it difficult to handle complex semantic understanding and reasoning tasks. Moreover, in an environment of dynamically changing data sources and updated knowledge bases, the real-time learning and adaptability of distance-supervised learning models need improvement.

Weakly-supervised learning is a method of training machine learning models using incomplete, inaccurate, or inconsistent labeled data. This approach often relies on heuristic rules, labeling functions, or the integration of multiple imperfect labeling sources. In NER, weakly-supervised learning can effectively utilize a large amount of unlabeled or partially labeled data and reduce reliance on manual labeling. This method enables the model to quickly adapt to new entity types and changing domains by integrating information from external knowledge bases, rules, or heuristic algorithms. It is especially suitable for professional or low-resource language scenarios. Although weakly-supervised learning reduces the requirement for large amounts of annotated data, data diversity and coverage must be ensured to avoid model bias and overfitting. At the same time, weakly-supervised data may contain some errors or inconsistent labels, requiring effective noise processing mechanisms, such as noise filtering or correction strategies. Fries et al. [150] used medical

ontology as the source of annotation heuristic rules and adopted a weakly-supervised learning method to train a medical entity classifier. Correct label noise by modeling the accuracy of each ontology and rules to improve model performance. Zhang et al. [151] improved the problems of insufficient label coverage and text noise by combining category description, keyword, and network structure analysis and using weakly-supervised learning methods with hierarchical structure information within the text. In addition, they also implemented a self-training strategy, which effectively enhanced the model's ability to handle complex, multi-faceted tasks, including improving the processing of labeled data. In order to solve the difficulties in weakly-supervised learning in existing NER, the following points may be the focus in the future. First, the model's adaptability in multiple languages and domains should be improved, and the data diversity brought by globalization should be responded to by developing cross-language and cross-domain transfer learning technologies. The second is to focus on data enhancement and simulation data generation and use technologies such as the generative adversarial network to make up for the lack of labeled data and enhance the robustness of the model. The third is to improve the processing method of noise labels and adopt a more accurate noise detection and correction mechanism to optimize the model's performance in complex data environments.

In addition to the above methods or strategies, other advanced technologies and methods have received increasing attention in NER research. For example, self-supervised learning utilizes large amounts of unlabeled data to learn useful feature representations. As an effective pre-training strategy, this method has been proven to significantly improve the performance of NER. In few-shot learning scenarios, meta-learning demonstrates its ability to quickly adapt to new tasks. By learning how to learn efficiently, meta-learning enables models to quickly adapt to new tasks or environments with limited data, which is especially important in fields where data is scarce. Additionally, incremental learning is another important approach that allows models to gradually adapt as they receive new data or face new tasks without completely retraining each time. This strategy is particularly effective in dealing with changing data environments because it keeps the model flexible and adaptable. Model compression and distillation techniques are becoming increasingly important in the NER field. These techniques reduce the size of large models, making them more suitable for environments with limited computing resources while maintaining or improving model performance. In summary, NER research is constantly evolving towards more efficient, adaptable, and resource-efficient directions, and these advanced methods and strategies manifest this trend.

7. Summary and prospect

This paper uses the literature in the field of NER obtained from the Web of Science core collection database as the data source. The following conclusions are drawn using CiteSpace software to comprehensively analyze NER's research status, existing achievements, research cooperation, research frontiers, and hotspots from macro and micro perspectives. The superior performance of deep learning in NER research has made the field of NER develop rapidly. According to the trend of the number of documents issued, the NER field is in a period of rapid development at this stage. From the perspective of research directions and journal distribution, NER research mainly involves computer science, medicine, biology, chemistry, and other disciplines, which shows that NER has interdisciplinary and cross-field common components. This multidisciplinary intersection has brought new application scenarios and research perspectives for developing NER technology, such as precise identification of biomedical named entities and entity identification of compound reactants. From the perspective of the cooperation between authors and the number of papers published, the core authors in the early stage of NER development include Munoz, R, Li, YP, and other authors, and the cooperation between those authors is close. In the mid-term, with Ananiadou, S, Xu, H, and other authors as the main body, the cooperation has become more intimate. Recent highly productive authors include Lin, HF, Qiu, QJ, and others. There are 63 prolific authors in the field, but the cooperative relationship between authors needs to be strengthened. This reminds us that deepening academic exchanges and cooperation not only contributes to knowledge sharing but also stimulates new creativity and technology integration, further promoting the innovation and development of NER research. From the perspective of the number of publications and cooperation among countries, countries with a higher volume of publications include PEOPLES R CHINA, the USA, ENGLAND, etc. The cooperation between countries is close. The number of publications in a country reflects, to a certain extent, the development level of NER technology in the language used in that country. The NER technology research in languages such as Chinese, English, and Arabic is significantly active. Meanwhile, we have also observed that other languages, such as Spanish, French, and German, are rapidly developing in NER technology. The cooperation between these countries shows the important role of international cooperation in promoting the global development of NER technology. In addition, the active research on multilingual NER technology reflects the urgent need to process multilingual information in the context of globalization. Encouraging international cooperation can accelerate technological progress and help promote information understanding and exchange in different languages and cultural backgrounds. From the perspective of inter-institutional cooperation and publication volume, the institutions with higher publication volume include Chinese Acad Sci, Harbin Inst Technol, and Dalian Univ Technol. The cooperation between institutions is mainly focused on the cooperation between universities, with less cooperation between schools and enterprises and less publication by enterprises. In the future, exploring and promoting cooperation models between universities and enterprises is expected to bring new opportunities for the application and industrialization of NER technology. The business community's demand for practical application of NER technology can provide rich practical scenarios for academic research. At the same time, the latest research results from academia can help companies solve technical challenges and rapidly transform and apply technology.

Based on the co-citation frequency of literature, the mainstream model BERT proposed by Devlin J. (2019) has made significant contributions to the development of NER. Lample G. (2016) paper extensively uses character-level information in NER tasks for the first time. This innovation provides new ideas for later processing of complex morphological languages (such as compound words in English). Vaswani A (2017) proposed the Transformer architecture, and its innovative attention mechanism marked an important

turning point in the field of NLP. In addition to the high frequency of co-cited literature mentioned above, other literature provides important methods and strategies for developing NER. The literature cluster analysis concludes that the research frontiers of NER include PLM, cross-language and cross-domain NER, nested and fine-grained NER, multimodal NER, few-shot NER, etc. Pre-trained language models such as BERT and its variants significantly improve the machine's ability to understand natural language by leveraging large amounts of text data to learn the deep features of the language. This progress not only brings a qualitative leap to the NER task but also provides new tools and methods for the entire field of natural language processing, especially in processing language context and understanding complex relationships. The progress of cross-language and cross-domain NER technology enables machines to better transfer and apply knowledge between different languages and fields, breaking down the barriers of language and professional knowledge and bringing new possibilities for global information sharing and knowledge management. Nested NER focuses on identifying mutually contained or overlapping entities, such as simultaneously annotating diseases and their related symptoms in the medical literature. Fine-grained NER strives to distinguish nuanced entity categories, such as further subdividing "organizations" into "non-profit organizations," "government agencies," etc. At the same time, fine-grained entity recognition and classification will promote the construction of a richer and more accurate knowledge map and provide basic support for developing the semantic web, intelligent search, and other technologies. The development of modal NER enables machines to more comprehensively understand and process data containing non-text information such as images and sounds, providing a new perspective for social media analysis, multimedia content management, and other applications. Research on a few-shot NER directly addresses the problem of data scarcity, enabling NER technology to quickly adapt to new fields or low-resource languages. With the in-depth development of these technologies, they are expected to have an important impact on intelligent search, personalized recommendation, intelligent assistants, and other fields. Although the pre-training model, cross-language and cross-domain NER, nested and fine-grained NER, multimodal NER, and small sample NER technologies have made significant progress in the field of NER in recent years, which has promoted the machine's ability to understand natural language and its application scope to expand significantly, especially in understanding language context, complex relationship processing, and multilingual information processing. However, there are still areas to be explored. Future research areas include but are not limited to: For the pre-trained model, how to further optimize its performance in processing long text, computational efficiency, and model interpretability, such as exploring innovative methods of model compression and interpretability mechanisms; In terms of cross language and cross domain NER, we will deeply study how to effectively deal with low resource languages with complex structure and changeable syntax through new in-depth learning methods, and how to better adapt and migrate the model to different professional fields, especially highly specialized fields; The main challenges for nested and fine-grained NER include accurately identifying and classifying intricately intertwined fine-grained entities in text, and understanding and parsing subtle relationships between entities; Multimodal NER, it is possible to explore and develop more diversified information fusion mechanisms in the future, focusing on how to make full use of and fuse information of multiple modes such as text, image, voice, etc. to improve the accuracy and robustness of entity recognition; In addition, the challenge of few-sample NER lies in how to use advanced technologies such as transfer learning and meta-learning to achieve rapid adaptation and improved generalization capabilities of the model under limited annotated data. The in-depth exploration of these research directions will not only fill the current technology gap but also greatly promote the development of NER technology in theory and practical applications, bringing new breakthroughs and innovations to the field of natural language processing.

From the keyword map analysis, the current research hotspots of NER include attention networks, multi-task joint learning, transfer learning, adversarial training, deep active learning, federated learning, distance-supervision learning, weakly-supervised learning, and other methods or strategies. These methods have greatly promoted the progress of models in understanding complex contexts and entity recognition accuracy. Among them, the attention mechanism enhances the model's focus on key information in the input data. It becomes a key factor in improving model performance, especially when understanding complex contexts and dealing with long-distance dependencies. Multi-task joint learning shows the potential to improve model generalization capabilities and learning efficiency by processing multiple related tasks in parallel in the same model. Transfer learning, especially the method of fine-tuning pre-trained models, shows excellent performance under low resource conditions and can significantly reduce the reliance on large amounts of labeled data. Adversarial training enhances the robustness of the model by introducing adversarial samples, helping the model maintain stable performance in the face of small perturbations in the input data. The deep active learning strategy effectively reduces the amount of annotation data needed by intelligently selecting samples with large amounts of information to annotate, which is especially suitable for scenarios with scarce annotation resources. Federated learning emphasizes jointly improving the model through collaborative training on multiple devices or servers while maintaining data privacy, which is particularly important for processing sensitive data. Distance supervision and weakly supervised learning effectively solve the problem of insufficient annotated data by utilizing existing knowledge bases or incompletely accurate annotated data to train models. However, these advances bring new possibilities to NER in practical applications. How to accurately transfer knowledge learned from one domain to another, handle differences in data distribution, and optimize models to adapt to new tasks remain challenges. In addition, applications in low-resource fields such as mechanical engineering and agricultural science face unique challenges, including but not limited to the accurate identification of professional terms and complex entities and the adaptability of models to domain-specific language patterns. For example, in the field of mechanical engineering, documents may be full of professional terms, technical parameters, CAD drawings, etc. The NER system is required not only to have high accuracy and robustness but also to be able to adapt to specific terms and expressions in various mechanical fields. Strategies such as adversarial training, deep active learning, federated learning, remote supervision, and weakly supervised learning show the potential to improve model robustness, reduce labeling requirements, and protect data privacy. However, how to effectively integrate these strategies to solve the specific problems in NER and improve the accuracy and efficiency of the model still needs further exploration.

Although this article attempts to analyze the NER field comprehensively, there are still certain limitations. First, during the

literature search process, “Named Entity Recognition” was selected as the primary search keyword, which ensured that we could effectively locate a wide range of literature directly related to NER. However, this search strategy may not fully cover all research literature in this field, especially those that may use different terms or keywords to describe similar concepts. In addition, interdisciplinary research or emerging technology applications may be published with different keywords, resulting in some unusual keywords not being included in this analysis. Future research can consider adopting more extensive search strategies, including more keywords and terms, to cover the literature in this field as comprehensively as possible.

Data availability

The data that was collected and analyzed during this study is contained in this published article and the data that was used to support the findings of this review are listed in the references at the end of the article.

CRedit authorship contribution statement

Jun Yang: Writing – original draft, Software, Formal analysis, Data curation, Conceptualization. **Taihua Zhang:** Writing – review & editing, Supervision. **Chieh-Yuan Tsai:** Writing – review & editing, Supervision. **Yao Lu:** Writing – review & editing, Supervision. **Liguo Yao:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Guizhou Provincial Basic Research Program (Natural Science) (Grant No. Qiankehejichu-ZK[2022] General 320), National Natural Science Foundation (Grant No. 72061006) and Academic New Seedling Foundation Project of Guizhou Normal University (Grant No. Qianshixinmiao-[2021]A30).

References

- [1] N.A. Chinchor, *Named entity task definition*, in: *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, 1995, pp. 317–332.
- [2] T.H. Dang, H.Q. Le, T.M. Nguyen, S.T. Vu, D3NER: biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information, *Bioinformatics* 34 (2018) 3539–3546, <https://doi.org/10.1093/bioinformatics/bty356>.
- [3] W.B. Karaa, E.H. Alkhamash, A. Bchir, Drug disease relation extraction from biomedical literature using NLP and machine learning, *Mobile Inf. Syst.* (2021) 2021, <https://doi.org/10.1155/2021/9958410>.
- [4] W. Hemati, A. Mehler, LSTMVoter: chemical named entity recognition using a conglomerate of sequence labeling tools, *J. Cheminf.* 11 (2019), <https://doi.org/10.1186/s13321-018-0327-2>.
- [5] A. Usie, R. Alves, F. Solsona, M. Vazquez, A. Valencia, CheNER: chemical named entity recognizer, *Bioinformatics* 30 (2014) 1039–1040, <https://doi.org/10.1093/bioinformatics/btt639>.
- [6] J.B. Antony, G.S. Mahalakshmi, *Content-based information retrieval by named entity recognition and verb semantic role labelling*, *J. Univers. Comput. Sci.* 21 (2015) 1830–1848.
- [7] M.E. Khademi, M. Fakhredanesh, Persian automatic text summarization based on named entity recognition, *Iranian Journal of Science and Technology-Transactions of Electrical Engineering* (2020), <https://doi.org/10.1007/s40998-020-00352-2>.
- [8] F.M. Guan, T. Tezuka, A medical Q&A system with entity linking and intent recognition, in: *2022 IEEE Symposium Series on Computational Intelligence (Ssci)*, 2022, pp. 820–829, <https://doi.org/10.1109/SSCI51031.2022.10022093>.
- [9] Z. Li, D. Qu, C.J. Xie, W.L. Zhang, Y.X. Li, Language model pre-training method in machine translation based on named entity recognition, *Int. J. Artif. Intell. Tool.* 29 (2020), <https://doi.org/10.1142/S0218213020400217>.
- [10] L. Wang, J.C. Jiang, J.W. Song, J. Liu, A weakly-supervised method for named entity recognition of agricultural knowledge graph, *Intelligent Automation and Soft Computing* 37 (2023) 833–848, <https://doi.org/10.32604/iase.2023.036402>.
- [11] D. Hanisch, K. Fundel, H.T. Mevissen, R. Zimmer, J. Fluck, ProMiner: rule-based protein and gene entity recognition, *BMC Bioinf.* 6 (2005), <https://doi.org/10.1186/1471-2105-6-S1-S14>.
- [12] A.P. Quimbaya, A.S. Munera, R.A.G. Rivera, J.C.D. Rodriguez, O.M.M. Velandia, A.A.G. Pena, C. Labbe, Named entity recognition over electronic health records through a combined dictionary-based approach, in: *International Conference on Enterprise Information Systems/International Conference on Project Management/International Conference on Health And Social Care Information Systems and Technologies, CENTERIS/PROJMAN/HICIST*, 2016, 2016, pp. 55–61, <https://doi.org/10.1016/j.procs.2016.09.123>.
- [13] P. McNamee, J. Mayfield, *Entity Extraction without Language-specific Resources*, 2002.
- [14] A. McCallum, W. Li, in: *Early Results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons*, 2003, pp. 188–191.
- [15] K. Humphreys, R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, Y. Wilks, University of Sheffield: Description of the LaSIE-II System as Used for MUC-7, 1998.
- [16] G.R. Krupka, K. Hausman, IsoQuest Inc.: Description of the NetOwl™ Extractor System as Used for MUC-7, 1998.
- [17] W.J. Black, F. Rinaldi, D. Mowatt, FACILE: Description of the NE System Used for MUC-7, 1998.
- [18] D.E. Appelt, J.R. Hobbs, J. Bear, D. Israel, M. Kameyama, A. Kehler, D. Martin, K. Myers, M. Tyson, SRI International FASTUS System MUC-6 Test Results and Analysis, 1995.
- [19] P. Liu, Y.M. Guo, F.L. Wang, G.H. Li, Chinese named entity recognition: the state of the art, *Neurocomputing* 473 (2022) 37–53, <https://doi.org/10.1016/j.neucom.2021.10.101>.
- [20] S.R. Eddy, Hidden Markov models, *Curr. Opin. Struct. Biol.* 6 (1996) 361–365, [https://doi.org/10.1016/S0959-440X\(96\)80056-X](https://doi.org/10.1016/S0959-440X(96)80056-X).
- [21] J.N. Kapur, *Maximum-entropy Models in Science and Engineering*, John Wiley & Sons, 1989.

- [22] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, *IEEE Intell. Syst. Their Appl.* 13 (1998) 18–28, <https://doi.org/10.1109/5254.708428>.
- [23] L. Rokach, O. Maimon, Top-down induction of decision trees classifiers - a survey, *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* 35 (2005) 476–487, <https://doi.org/10.1109/TSMCC.2004.843247>.
- [24] J. Lafferty, A. McCallum, F.C.N. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, 2001.
- [25] B.G. Mi, Y. Fan, A review: development of named entity recognition (NER) technology for aeronautical information intelligence, *Artif. Intell. Rev.* 56 (2023) 1515–1542, <https://doi.org/10.1007/s10462-022-10197-2>.
- [26] M. Marrero, J. Urbano, S. Sanchez-Cuadrado, J. Morato, J.M. Gomez-Berbis, Named entity recognition: fallacies, challenges and opportunities, *Comput. Stand. Interfac.* 35 (2013) 482–489, <https://doi.org/10.1016/j.csi.2012.09.004>.
- [27] A. Goyal, V. Gupta, M. Kumar, Recent named entity recognition and classification techniques: a systematic review, *Computer Science Review* 29 (2018) 21–43, <https://doi.org/10.1016/j.cosrev.2018.06.001>.
- [28] Z. Nasar, S.W. Jaffry, M.K. Malik, Named entity recognition and relation extraction: state-of-the-art, *ACM Comput. Surv.* 54 (2021), <https://doi.org/10.1145/3445965>.
- [29] J. Li, A.X. Sun, J.L. Han, C.L. Li, A survey on deep learning for named entity recognition, *IEEE Trans. Knowl. Data Eng.* 34 (2022) 50–70, <https://doi.org/10.1109/TKDE.2020.2981314>.
- [30] Y.X. Li, Y. Wang, X. Rui, Y.X. Li, Y. Li, H.Z. Wang, J. Zuo, Y.D. Tong, Sources of atmospheric pollution: a bibliometric analysis, *Scientometrics* 112 (2017) 1025–1045, <https://doi.org/10.1007/s11192-017-2421-z>.
- [31] D.J. Yu, T.X. Pan, Tracing knowledge diffusion of TOPSIS: a historical perspective from citation network, *Expert Syst. Appl.* 168 (2021), <https://doi.org/10.1016/j.eswa.2020.114238>.
- [32] D.J. Yu, L.B. Sheng, Z.S. Xu, Analysis of evolutionary process in intuitionistic fuzzy set theory: a dynamic perspective, *Inf. Sci.* 601 (2022) 175–188, <https://doi.org/10.1016/j.ins.2022.04.019>.
- [33] C. Chen, L. Leydesdorff, *Technology, Patterns of Connections and Movements in Dual-map Overlays: A New Method of Publication Portfolio Analysis*, 65 (2014) 334–351.
- [34] C.M. Chen, Z.G. Hu, S.B. Liu, H. Tseng, Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace, *Expet Opin. Biol. Ther.* 12 (2012) 593–608, <https://doi.org/10.1517/14712598.2012.674507>.
- [35] G.R. Doddington, A. Mitchell, M.A. Przybocki, L.A. Ramshaw, S. Strassel, R.M. Weischedel, *The automatic content extraction (ACE) program – tasks, data, and evaluation*, in: *International Conference on Language Resources and Evaluation*, 2004.
- [36] J.R. Cao, E.M. van Veen, N. Peek, A.G. Renehan, S. Ananiadou, A novel automated approach to mutation-cancer relation extraction by incorporating heterogeneous knowledge, 27, in: *Ieee Journal of Biomedical and Health Informatics*, 2023, pp. 1096–1105, <https://doi.org/10.1109/JBHI.2022.3220924>.
- [37] K. Espinosa, P. Georgiadis, F. Christopoulou, M.Z. Ju, M. Miwa, S. Ananiadou, Comparing neural models for nested and overlapping biomedical event detection, *BMC Bioinf.* 23 (2022), <https://doi.org/10.1186/s12859-022-04746-3>.
- [38] Q. Wei, Y.Y. Zhang, M. Amith, R. Lin, J. Lapeyrolerie, C. Tao, H. Xu, Recognizing software names in biomedical literature using machine learning, *Health Inf. J.* 26 (2020) 21–33, <https://doi.org/10.1177/1460458219869490>.
- [39] M. Jiang, Y.K. Chen, M. Liu, S.T. Rosenbloom, S. Mani, J.C. Denny, H. Xu, A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries, *J. Am. Med. Inf. Assoc.* 18 (2011) 601–606, <https://doi.org/10.1136/amiajnl-2011-000163>.
- [40] T.X. Zhang, H.F. Lin, Y.Q. Ren, Z.H. Yang, J. Wang, S.W. Zhang, B. Xu, X.D. Duan, Adversarial transfer network with bilinear attention for the detection of adverse drug reactions from social media, *Appl. Soft Comput.* 106 (2021), <https://doi.org/10.1016/j.asoc.2021.107358>.
- [41] T.X. Zhang, H.F. Lin, Y.Q. Ren, Z.H. Yang, J. Wang, X.D. Duan, B. Xu, Identifying adverse drug reaction entities from social media with adversarial transfer learning model, *Neurocomputing* 453 (2021) 254–262, <https://doi.org/10.1016/j.neucom.2021.05.007>.
- [42] X. Lv, Z. Xie, D.X. Xu, X.G. Jin, K. Ma, L.F. Tao, Q.J. Qiu, Y.S. Pan, Chinese named entity recognition in the geoscience domain based on BERT, *Earth Space Sci.* 9 (2022), <https://doi.org/10.1029/2021EA002166>.
- [43] Q.J. Qiu, Z. Xie, L. Wu, L.F. Tao, W.J. Li, BiLSTM-CRF for geological named entity recognition from the geoscience literature, *EARTH SCIENCE INFORMATICS* 12 (2019) 565–579, <https://doi.org/10.1007/s12145-019-00390-3>.
- [44] D.J. Price, *Solla de*, in: *Little Science, Big Science*, Columbia University Press, New York, 1963, 1963.
- [45] D.J. Yu, G. Kou, Z.S. Xu, S.S. Shi, Analysis of collaboration evolution in AHP research: 1982-2018, *Int. J. Inf. Technol. Decis. Making* 20 (2021) 7–36, <https://doi.org/10.1142/S0219622020500406>.
- [46] G. Qian, Scientometric sorting by importance for literatures on life cycle assessments and some related methodological discussions, *Int. J. Life Cycle Assess.* 19 (2014) 1462–1467, <https://doi.org/10.1007/s11367-014-0747-9>.
- [47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, vol. 30, NIPS, 2017.), 2017.
- [48] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, in: *North American Chapter of the Association for Computational Linguistics*, 2016.
- [49] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics* 36 (2020) 1234–1240, <https://doi.org/10.1093/bioinformatics/btz682>.
- [50] R. Leaman, C.H. Wei, Z.Y. Lu, tmChem: a high performance approach for chemical named entity recognition and normalization, *J. Cheminf.* 7 (2015), <https://doi.org/10.1186/1758-2946-7-S1-S3>.
- [51] E.F. Sang, F.J.a.p.c. De Meulder, *Introduction to the CoNLL-2003 Shared Task: Language-independent Named Entity Recognition*, 2003.
- [52] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, *Natural Language Processing (Almost) from Scratch*, vol. 12, 2011, pp. 2493–2537.
- [53] A.M. Dai, Q.V. Le, *Semi-supervised Sequence Learning*, 2015, p. 28.
- [54] J.P.C. Chiu, E. Nichols, Named entity recognition with bidirectional LSTM-CNNs, *Transactions of the association for computational linguistics* 4 (2016) 357–370.
- [55] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, *Association for Computational Linguistics* (2018) 2227–2237, <https://doi.org/10.18653/v1/N18-1202>.
- [56] D.J. Yu, Z.P. Yan, Combining machine learning and main path analysis to identify research front: from the perspective of science-technology linkage, *Scientometrics* 127 (2022) 4251–4274, <https://doi.org/10.1007/s11192-022-04443-1>.
- [57] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, L. Assoc Computat, BERT: pre-training of deep bidirectional transformers for language understanding, in: *2019 Conference of The North American Chapter of The Association for Computational Linguistics: Human Language Technologies (Naacl Hlt 2019)*, vol. 1, 2019, pp. 4171–4186.
- [58] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: a robustly optimized bert pretraining approach. *arXiv Preprint arXiv:1907.11692*, 2019.
- [59] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: a lite bert for self-supervised learning of language representations, *arXiv preprint arXiv:1909.1194* (2019).
- [60] A. Agrawal, S. Tripathi, M. Vardhan, V. Sihag, G. Choudhary, N. Dragoni, BERT-based transfer-learning approach for nested named-entity recognition using joint labeling, *APPLIED SCIENCES-BASEL* 12 (2022), <https://doi.org/10.3390/app12030976>.
- [61] M.J. Chen, X. Luo, H.L. Shen, Z.Y. Huang, Q.J. Peng, A novel named entity recognition scheme for steel E-commerce platforms using a lite BERT, *CMES-COMPUTER MODELING IN ENGINEERING & SCIENCES* 129 (2021) 47–63, <https://doi.org/10.32604/cmcs.2021.017491>.
- [62] X.Y. Li, H. Zhang, X.H. Zhou, Chinese clinical named entity recognition with variant neural structures based on BERT methods, *J. Biomed. Inf.* (2020) 107, <https://doi.org/10.1016/j.jbi.2020.103422>.
- [63] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *Improving Language Understanding by Generative Pre-training*, 2018.

- [64] Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal S., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever L., Amodei D., Language Models Are Few-Shot Learners, vol. 33, 2020, pp. 1877–1901.
- [65] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R.R. Salakhutdinov, Q.V.J.A.i.n.i.p.s. Le, Xlnet: Generalized Autoregressive Pretraining for Language Understanding, 2019, p. 32.
- [66] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q.V. Le, R.J.a.p.a. Salakhutdinov, Transformer-xl: Attentive Language Models beyond a Fixed-Length Context, 2019.
- [67] Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, H.J.a.p.a. Wu, Ernie: Enhanced Representation through Knowledge Integration, 2019.
- [68] K. Clark, M.-T. Luong, Q.V. Le, C.D. Manning, Electra: Pre-training Text Encoders as Discriminators rather than Generators, 2020.
- [69] S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, G.J.a.p.a. Wang, Gpt-ner: Named Entity Recognition via Large Language Models, 2023.
- [70] E.J.a.p.a. Covas, Named Entity Recognition Using GPT for Identifying Comparable Companies, 2023.
- [71] R. Yan, X. Jiang, D.J.N.P.L. Dang, Named Entity Recognition by Using XLNet-BiLSTM-CRF, vol. 53, 2021, pp. 3339–3356.
- [72] D. Yang, F. Wan, Y. Zhang, Named entity recognition in XLNet cyberspace security domain based on dictionary embedding, in: 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC), IEEE, 2022, pp. 1–5.
- [73] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V.J.a.p.a. Stoyanov, Unsupervised Cross-Lingual Representation Learning at Scale, 2019.
- [74] P. Keung, Y. Lu, V.J.a.p.a. Bhardwaj, Adversarial Learning with Contextual Embeddings for Zero-Resource Cross-Lingual Classification and NER, 2019.
- [75] X. Feng, X. Feng, B. Qin, Z. Feng, T. Liu, Improving Low Resource Named Entity Recognition Using Cross-Lingual Knowledge Transfer, IJCAI, 2018, pp. 4071–4077.
- [76] Z. Liu, Y. Xu, T. Yu, W. Dai, Z. Ji, S. Cahyawijaya, A. Madotto, P. Fung, Crossner: evaluating cross-domain named entity recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 13452–13460.
- [77] C. Jia, X. Liang, Y. Zhang, Cross-domain NER using cross-domain language modeling, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2464–2474.
- [78] S. Chen, G. Aguilar, L. Neves, T.J.a.p.a. Solorio, Data Augmentation for Cross-Domain Named Entity Recognition, 2021.
- [79] A. Brack, A. Hoppe, P. Buschermöhle, R. Ewerth, Cross-domain multi-task learning for sequential sentence classification in research papers, in: Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries, 2022, pp. 1–13.
- [80] Z.H. Liu, Y. Xu, T.Z. Yu, W.L. Dai, Z.W. Ji, S. Cahyawijaya, A. Madotto, P. Fung, I. Assoc, Advancement artificial, CrossNER: evaluating cross-domain named entity recognition, in: Thirty-Fifth Aaai Conference On Artificial Intelligence, Thirty-Third Conference On Innovative Applications Of Artificial Intelligence And The Eleventh Symposium On Educational Advances In Artificial Intelligence, 2021, pp. 13452–13460.
- [81] Q. Peng, C.M. Zheng, Y. Cai, T. Wang, H.R. Xie, Q. Li, Unsupervised cross-domain named entity recognition using entity-aware adversarial training, Neural Network. 138 (2021) 68–77, <https://doi.org/10.1016/j.neunet.2020.12.027>.
- [82] A. Katiyar, C. Cardie, Nested named entity recognition revisited, in: North American Chapter of the, Association for Computational Linguistics, 2018.
- [83] Y. Wang, H.H. Tong, Z.Y. Zhu, Y. Li, Nested named entity recognition: a survey, ACM Trans. Knowl. Discov. Data 16 (2022), <https://doi.org/10.1145/3522593>.
- [84] D. Shen, J. Zhang, G. Zhou, J. Su, C.L. Tan, Effective adaptation of hidden Markov model-based named entity recognizer for biomedical domain, in: BioNLP@ ACL, 2003.
- [85] M. Ju, M. Miwa, S. Ananiadou, A neural layered model for nested named entity recognition, in: North American Chapter of the Association for Computational Linguistics, 2018.
- [86] M.B. Xu, H. Jiang, S. Watcharawittayakul, A local detection approach for named entity recognition and mention detection, in: Proceedings of The 55th Annual Meeting Of The Association For Computational Linguistics (Acl 2017), vol. 1, 2017, pp. 1237–1247, <https://doi.org/10.18653/v1/P17-1114>.
- [87] W. Lu, D. Roth, Joint mention extraction and classification with mention hypergraphs, in: Conference On Empirical Methods in Natural Language Processing, 2015.
- [88] B.L. Wang, W. Lu, Y. Wang, H.X. Jin, L. Assoc Computat, A neural transition-based model for nested mention recognition, in: 2018 Conference On Empirical Methods in Natural Language Processing (Emnlp 2018), 2018, pp. 1011–1017.
- [89] R.S. Geng, Y.P. Chen, R.Z. Huang, Y.B. Qin, Q.H. Zheng, Planarized sentence representation for nested named entity recognition, Inf. Process. Manag. 60 (2023), <https://doi.org/10.1016/j.ipm.2023.103352>.
- [90] S.M. Cui, I. Joe, A Multi-Head Adjacent Attention-Based Pyramid Layered Model for Nested Named Entity Recognition, vol. 35, Neural Computing & Applications, 2023, pp. 2561–2574, <https://doi.org/10.1007/s00521-022-07747-8>.
- [91] Y.P. Chen, R. Huang, L.J. Pan, R.Z. Huang, Q.H. Zheng, P. Chen, A controlled attention for nested named entity recognition, Cognitive Computation 15 (2023) 132–145, <https://doi.org/10.1007/s12559-023-10112-z>.
- [92] A.J.C. Rodríguez, D.C. Castro, S.H. García, Noun-based Attention Mechanism for Fine-Grained Named Entity Recognition, vol. 193, 2022 116406.
- [93] Q. Wan, L. Wei, S. Zhao, J.J.K.-B.S. Liu, A span-based multi-modal attention network for joint entity-relation extraction 262 (2023) 110228.
- [94] X. Wang, V. Hu, X. Song, S. Garg, J. Xiao, J. Han, ChemNER: fine-grained chemistry named entity recognition with ontology-guided distant supervision, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021.
- [95] J. Yu, J. Jiang, L. Yang, R. Xia, in: Improving Multimodal Named Entity Recognition via Entity Span Detection with Unified Multimodal Transformer, Association for Computational Linguistics, 2020.
- [96] D. Zhang, S. Wei, S. Li, H. Wu, Q. Zhu, G. Zhou, Multi-modal graph fusion for named entity recognition with targeted visual guidance, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 14347–14355.
- [97] B. Zhou, Y. Zhang, K. Song, W. Guo, G. Zhao, H. Wang, X. Yuan, A span-based multimodal variational autoencoder for semi-supervised multimodal named entity recognition, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 6293–6302.
- [98] H. Wang, L. Cheng, W. Zhang, D.W. Soh, L.J.a.p.a. Bing, Enhancing Few-Shot NER with Prompt Ordering Based Data Augmentation, 2023.
- [99] J. Chen, Q. Liu, H. Lin, X. Han, L.J.a.p.a. Sun, Few-shot Named Entity Recognition with Self-Describing Networks, 2022.
- [100] S.S.S. Das, A. Katiyar, R.J. Passonneau, R.J.a.p.a. Zhang, CONTAiNER: Few-Shot Named Entity Recognition via Contrastive Learning, 2021.
- [101] Y. Chen, Y. Zheng, Z.J.a.p.a. Yang, Prompt-Based Metric Learning for Few-Shot NER, 2022.
- [102] D.-H. Lee, A. Kadakia, K. Tan, M. Agarwal, X. Feng, T. Shibuya, R. Mitani, T. Sekiya, J. Pujara, X.J.a.p.a. Ren, Good Examples Make a Faster Learner: Simple Demonstration-Based Learning for Low-Resource NER, 2021.
- [103] Y. Shen, Z. Tan, S. Wu, W. Zhang, R. Zhang, Y. Xi, W. Lu, Y.J.a.p.a. Zhuang, PromptNER: Prompt Locating and Typing for Named Entity Recognition, 2023.
- [104] D.J. Yu, Z.S. Xu, W. Pedrycz, W.R. Wang, Information sciences 1968-2016: a retrospective analysis with text mining and bibliometric, Inf. Sci. 418 (2017) 619–634, <https://doi.org/10.1016/j.ins.2017.08.031>.
- [105] J.I. Kazama, T. Makino, Y. Ohta, J. Tsujii, Tuning support vector machines for biomedical named entity recognition, in: ACL Workshop on Natural Language Processing in the Biomedical Domain, 2002.
- [106] T. Rocktaschel, M. Weidlich, U. Leser, ChemSpot: a hybrid system for chemical named entity recognition, Bioinformatics 28 (2012) 1633–1640, <https://doi.org/10.1093/bioinformatics/bts183>.
- [107] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space, 2013 arXiv preprint arXiv:1301.3781.
- [108] Huang Z., Xu W., Yu K., Bidirectional LSTM-CRF models for sequence tagging, arXiv preprint arXiv:1508.01991, 2015.
- [109] C. Fraley, A.E. Raftery, Model-based clustering, discriminant analysis, and density estimation, J. Am. Stat. Assoc. 97 (2002) 611–631, <https://doi.org/10.1198/016214502760047131>.
- [110] D.J. Yu, Y. Liu, Z.S. Xu, Analysis of knowledge evolution in PROMETHEE: a longitudinal and dynamic perspective, Inf. Sci. 642 (2023), <https://doi.org/10.1016/j.ins.2023.119151>.

- [111] J.C. Donohue, Understanding Scientific Literatures: A Bibliometric Approach, 1973.
- [112] H.F. He, X. Sun, A unified model for cross-domain and semi-supervised named entity recognition in Chinese social media, in: THIRTY-FIRST AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2017, pp. 3216–3222.
- [113] Z.Y. Niu, G.Q. Zhong, H. Yu, A review on the attention mechanism of deep learning, *Neurocomputing* 452 (2021) 48–62, <https://doi.org/10.1016/j.neucom.2021.03.091>.
- [114] D. Bahdanau, K. Cho, Y.J.C. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, 2014 abs/1409.0473.
- [115] K. Xu, J.L. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R.S. Zemel, Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, *INTERNATIONAL CONFERENCE ON MACHINE LEARNING* 37 (2015) 2048–2057.
- [116] S. Zhang, Y. Sheng, J.F. Gao, J.H. Chen, J.J. Huang, S.F. Lin, A multi-domain named entity recognition method based on part-of-speech attention mechanism, in: COMPUTER SUPPORTED COOPERATIVE WORK AND SOCIAL COMPUTING, CHINESECSW, 2019, 2019, pp. 631–644, https://doi.org/10.1007/978-981-15-1377-0_49.
- [117] J.C.W. Lin, Y.N. Shao, Y. Djenouri, U. Yun, ASRNN: a recurrent neural network with an attention model for sequence labeling, *Knowl. Base Syst.* 212 (2021), <https://doi.org/10.1016/j.knosys.2020.106548>.
- [118] G.H. Xu, C.Y. Wang, X.F. He, Improving clinical named entity recognition with global neural attention, *WEB AND BIG DATA (APWEB-WAIM 2018)*, PT II (2018) 264–279, https://doi.org/10.1007/978-3-319-96893-3_20.
- [119] C. Zhuang, X. Jin, W. Zhu, J.W. Liu, L. Bai, X.Q. Cheng, Deep learning based relation extraction: a survey, *Chinese Journal of Informatics* 33 (2019) 1–18.
- [120] Z.Q. Geng, Y.H. Zhang, Y.M. Han, Joint entity and relation extraction model based on rich semantics, *Neurocomputing* 429 (2021) 132–140, <https://doi.org/10.1016/j.neucom.2020.12.037>.
- [121] Z.Y. Zhang, X.B. Shu, T.W. Liu, Z. Fang, Q.G. Li, Ieee, joint entity linking and relation extraction with neural networks for knowledge base population, in: 2020 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2020.
- [122] Q. Li, H. Ji, Incremental joint extraction of entity mentions and relations, in: PROCEEDINGS OF THE 52ND ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, vol. 1, 2014, pp. 402–412, <https://doi.org/10.3115/v1/p14-1038>.
- [123] K. Zhao, H. Xu, Y. Cheng, X. Li, K.J.K.-B.S. Gao, Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction 219 (2021) 106888.
- [124] T. Chen, L. Zhou, N. Wang, X.J.A.S.C. Chen, Joint Entity and Relation Extraction with Position-Aware Attention and Relation Embedding, vol. 119, 2022 108604.
- [125] Z. Yan, S. Yang, W. Liu, K.J.a.p.a. Tu, Joint Entity and Relation Extraction with Span Pruning and Hypergraph Neural Networks, 2023.
- [126] G. Bekoulis, J. Deleu, T. Demeester, C. Develder, Joint entity recognition and relation extraction as a multi-head selection problem 114 (2018) 34–45.
- [127] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, H.J.a.p.a. Hajishirzi, A General Framework for Information Extraction Using Dynamic Span Graphs, 2019.
- [128] D.Q. Nguyen, PhoNLP: a joint multi-task learning model for Vietnamese part-of-speech tagging, named entity recognition and dependency parsing, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, 2021, pp. 1–7.
- [129] S.J. Pan, Q.A. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359, <https://doi.org/10.1109/TKDE.2009.191>.
- [130] D. Peng, Y. Wang, C. Liu, Z.J.I.S.F. Chen, TL-NER, A Transfer Learning Model for Chinese Named Entity Recognition, vol. 22, 2020, pp. 1291–1304.
- [131] L. Gligic, A. Kormilitzin, P. Goldberg, A.J.N.N. Nevado-Holgado, Named entity recognition in electronic health records using transfer learning bootstrapped, *Neural Network* 121 (2020) 132–139.
- [132] Y.Q. Yu, Y.Z. Wang, J.Q. Mua, W. Li, S.T. Jiao, Z. Wang, P. Lv, Y.Q. Zhu, Chinese mineral named entity recognition based on BERT model, *Expert Syst. Appl.* 206 (2022), <https://doi.org/10.1016/j.eswa.2022.117727>.
- [133] L.G. Yao, H.S. Huang, K.W. Wang, S.H. Chen, Q.Q. Xiong, Fine-grained mechanical Chinese named entity recognition based on ALBERT-AttBiLSTM-CRF and transfer learning, *SYMMETRY-BASEL* 12 (2020), <https://doi.org/10.3390/sym12121986>.
- [134] I.J. Goodfellow, J. Shlens, C.J.a.p.a. Szegedy, Explaining and Harnessing Adversarial Examples, 2014.
- [135] G.H. Wen, H.H. Chen, H.H. Li, Y. Hu, Y.H. Li, C.J. Wang, Cross domains adversarial learning for Chinese named entity recognition for online medical consultation, *J. Biomed. Inf.* (2020) 112, <https://doi.org/10.1016/j.jbi.2020.103608>.
- [136] C. Park, S. Jeong, J. Kim, ADMit: Improving NER in Automotive Domain with Domain Adversarial Training and Multi-Task Learning, vol. 225, 2023 120007.
- [137] J. Wang, W. Xu, X. Fu, G. Xu, Y.J.K.-B.S. Wu, ASTRAL: Adversarial Trained LSTM-CNN for Named Entity Recognition, vol. 197, 2020 105842.
- [138] P.Z. Ren, Y. Xiao, X.J. Chang, P.Y. Huang, Z.H. Li, B.B. Gupta, X.J. Chen, X. Wang, A survey of deep active learning, *ACM Comput. Surv.* 54 (2022), <https://doi.org/10.1145/3472291>.
- [139] S. Tong, in: *Active Learning: Theory and Applications*, 2001.
- [140] A. Agrawal, S. Tripathi, M. Vardhan, Active Learning Approach Using a Modified Least Confidence Sampling Strategy for Named Entity Recognition, vol. 10, 2021, pp. 113–128.
- [141] W. Li, Y.J. Du, X.Y. Li, X.L. Chen, C.Z. Xie, H. Li, X.L. Li, UD-BBC: named entity recognition in social network combined BERT-BiLSTM-CRF with active learning, *Eng. Appl. Artif. Intell.* (2022) 116, <https://doi.org/10.1016/j.engappai.2022.105460>.
- [142] P. Radmard, Y. Fathullah, A. Lipani, Subsequence based deep active learning for named entity recognition, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 4310–4321.
- [143] J. Konečný, H.B. McMahan, F.X. Yu, P. Richtárik, A.T. Suresh, D.J.a.p.a. Bacon, Federated Learning: Strategies for Improving Communication Efficiency, 2016.
- [144] C. Wu, F. Wu, L. Lyu, Y. Huang, X.J.N.c. Xie, Communication-efficient Federated Learning via Knowledge Distillation, vol. 13, 2022, p. 2032.
- [145] R. Wang, T. Yu, J. Wu, H. Zhao, S. Kim, R. Zhang, S. Mitra, R. Henao, Federated domain adaptation for named entity recognition via distilling with heterogeneous tag sets, in: Findings of the Association for Computational Linguistics vol. 2023, ACL, 2023, pp. 7449–7463.
- [146] S. Ge, F. Wu, C. Wu, T. Qi, Y. Huang, X.J.a.p.a. Xie, Fedner: Privacy-Preserving Medical Named Entity Recognition with Federated Learning, 2020.
- [147] Q. Li, T. Xie, P. Peng, H. Wang, G. Wang, A class-rebalancing self-training framework for distantly-supervised named entity recognition, in: Findings of the Association for Computational Linguistics vol. 2023, ACL, 2023, pp. 11054–11068.
- [148] K. Zhou, Y. Li, Q.J.a.p.a. Li, Distantly Supervised Named Entity Recognition via Confidence-Based Multi-Class Positive and Unlabeled Learning, 2022.
- [149] Y. Meng, Y. Zhang, J. Huang, X. Wang, Y. Zhang, H. Ji, J.J.a.p.a. Han, Distantly-supervised Named Entity Recognition with Noise-Robust Learning and Language Model Augmented Self-Training, 2021.
- [150] J.A. Fries, E. Steinberg, S. Khattar, S.L. Fleming, J. Posada, A. Callahan, N.H.J.N.c. Shah, Ontology-driven Weak Supervision for Clinical Entity Classification in Electronic Health Records, vol. 12, 2021, p. 2017.
- [151] Y. Zhang, B. Jin, X. Chen, Y. Shen, Y. Zhang, Y. Meng, J.J.a.p.a. Han, Weakly Supervised Multi-Label Classification of Full-Text Scientific Papers, 2023.