Journal of Infrastructure, Policy and Development 2024, 8(15), 9789.
https://doi.org/10.24294/jipd9789

Review

Named entity recognition in government domain: A systematic literature review

Tosan Wiar Ramdhani*, **Indra Budi, Betty Purwandari**

Universitas Indonesia, Depok 16424, Indonesia
*** Corresponding author:** Tosan Wiar Ramdhani, tosan.wiar01@ui.ac.id

CITATION

Ramdhani TW, Budi I, Purwandari B. (2024). Named entity recognition in government domain: A systematic literature review. Journal of Infrastructure, Policy and Development. 8(15): 9789. https://doi.org/10.24294/jipd9789

ARTICLE INFO

Received: 22 October 2024
Accepted: 20 November 2024
Available online: 10 December 2024

Abstract: Named Entity Recognition (NER), a core task in Information Extraction (IE) alongside Relation Extraction (RE), identifies and extracts entities like place and person names in various domains. NER has improved business processes in both public and private sectors but remains underutilized in government institutions, especially in developing countries like Indonesia. This study examines which government fields have utilized NER over the past five years, evaluates system performance, identifies common methods, highlights countries with significant adoption, and outlines current challenges. Over 64 international studies from 15 countries were selected using PRISMA 2020 guidelines. The findings are synthesized into a preliminary ontology design for Government NER.

Keywords: named entity recognition; information extraction; machine learning; deep learning; government domain; natural language processing

1. Introduction

Named Entity Recognition (NER) is one of the primary tasks of Information Extraction (IE), in addition to Relation Extraction (RE), which aims to extract entities such as place names or individuals for general domains or specific entities for specific domains (Gasmi et al., 2019). People, locations, and organizations comprise the most prevalent generic domain NER entities. Conversely, for specific domains, it may be a specific entity, such as the names of DNA or proteins, for Biomedical NER domains.

NER was first introduced at the sixth Message Understanding Conference (MUC-6) to help advance the field of information extraction in 1995 (Mansouri et al., 2008). The first two entities extracted at that Conference are person and organization (Grishman and Sundheim., 1995). NER has been employed in a variety of domains, including biomedical, legal, and government, as well as in generic domains.

NER has assisted both the public and private sectors in the improvement of certain business processes within an organization. NER can facilitate the talent management process in the human resources domain by extracting human resources-related entities for employee allocation (Cenikj et al.,2021). NER can assist the Italian local government in automating the electronic invoicing process by extracting finance entities, including quantity, unit cost, discount, and tax, in the government domain (Pugliese et al., 2021).

In terms of methods, there are four mainstream approaches in NER: i) rule-based approaches, which do not require annotated datasets but require some hand-crafted rules to extract entities; ii) unsupervised learning approaches, which rely on unsupervised algorithms using unlabeled datasets; iii) supervised learning approaches, which use supervised algorithms and feature engineering to assist in entity extraction;

and iv) deep learning-based approaches, which do not require any feature engineering but can automatically extract entities by training several labeled datasets using a deep neural network model (Li et al., 2022). Each approach to the NER method has its advantages and disadvantages. For instance, rule-based NER is highly efficient when the dataset is semi-structured or similar, necessitating fewer hand-crafted rules. However, an extensive collection of hand-crafted rules will be produced when the dataset is diverse and has a complex structure. On the other hand, deep learning NER may offer the advantage of not requiring any features engineering to generate the NER model. Nevertheless, the dataset must frequently be manually labeled, and the training time is contingent upon the size of the dataset and the architecture of the deep neural network. Since unsupervised learning methods are relatively rare, especially in the government domain, we only tried to compare rule-based, supervised learning approaches, which mostly are machine learning, deep learning, and combined methods.

In terms of performance measurement, generic domain NER has reached a high number of 96.76% F1 score by using the Bidirectional Encoder Representations from Transformers (BERT) model on Conference on a Computational Natural Language Learning (CoNLL) 03 dataset (Devlin et al., 2019). On the other hand, the efficacy of specific domain NER can differ based on the domain, languages, datasets, or methods. The NER efficacy for Russian strategic documents can be as low as 57% in specific domains, such as government (Ivanin et al., 2021). Nevertheless, it may reach as high as 99.4% for government policy purposes (Sufi, 2022). NER is a potent Natural Language Processing (NLP) technique that has the potential to enhance the efficacy and effectiveness of government institutions.

The implementation of NER in government institutions has been advantageous to numerous countries in a variety of sectors. For example, in Italy, research has implemented NER to extract electronic invoices that adhere to a standard electronic format established by the local government (Pugliese et al., 2021). This is merely one illustration of the applications of NER; comparable applications have been investigated in other countries. For instance, in China, research has employed NER to extract entities from government procurement documents, thereby enabling the development of a government procurement knowledge graph for public use. Furthermore, the Bogota city administration in Colombia has implemented NER to identify traffic accidents by extracting place entities from Twitter (Suat-Rojas et al., 2022). These cases demonstrate the diverse applications of NER in various government sectors.

This study was motivated by the diverse results of both performance and methods in specific domains. The objective is to conduct a survey on NER in the government domain, which will investigate the fields in which NER has assisted government institutions over the past seven years, the performance of NER in the government domain, the methods or approaches that have been most commonly used, the countries in which NER is being used most frequently, and the open problems and challenges of NER in the government domain.

Contribution of this review. Another motivation for conducting this study is the need to explore state-of-the-art approaches in the NER government domain, which is an essential part of the author's dissertation study. The SLR also serves as a foundation for evaluating and comparing the efficacy of the proposed strategies in the NER governance domain. The term "government domain" in NER research denotes research that employs NER methods on government data or research that has a government-related impact.

2. Materials and methods

This article provides an SLR of NER research conducted within the government domain, adhering to the established PRISMA 2020 guidelines (PRISMA, 2020). SLR is a research methodology that emphasizes the systematic collection and comprehensive evaluation of current research on a specific topic, resulting in the synthesis of well-informed conclusions that are based on a broader body of knowledge.

PRISMA 2020 has been used as a guideline to review information technology government-related studies (Sensue et al., 2022). In the context of our study, the central theme revolves around NER in the government domain. The PRISMA 2020 checklist has been instrumental in guiding the process of this SLR. However, it is important to acknowledge that our study has prudently implemented 18 of the 27 criteria items that were specified. It is important to note that the remaining nine checklist items were purposefully omitted, with the majority of these exclusions being associated with the assessment of certainty and the risk of bias in the study. To guarantee that our SLR method delivers the most pertinent and comprehensive analysis, these strategic exclusions are predicated on the distinctive attributes of the NER research landscape in the government domain.

The decision to eliminate the risk of study bias and assessment of certainty in our systematic literature review, which was conducted collaboratively by our team of authors, was primarily motivated by the distinctive characteristics of the data we collected during our extensive research. Our comprehensive search efforts in the government sector yielded a rather limited corpus of 64 scholarly articles within the realm of Named Legal Entity Recognition (NER). What is particularly intriguing about these articles is that they each present datasets customized to their respective research objectives, thereby introducing a high degree of diversity and uniqueness among the data sources. This lack of standardization and comparability in our datasets has resulted in a more complex application of traditional risk of bias assessment and certainty evaluation due to the uniqueness of the datasets. Furthermore, the multitude of data sources is frequently accompanied by a need for standardized reporting conventions and varying levels of data detail. This factor significantly impedes the conventional sense of conducting robust assessments. Conducting an exhaustive evaluation of the risk of bias and certainty within the context of a uniformly structured dataset has yet to be rendered possible due to the complexities of the situation, which include the use of multiple data sources and methodological approaches. Consequently, we have decided to confront this complexity by recognizing the inherent constraints of data sets and focusing our efforts on the presentation of a qualitative synthesis of the findings, a comprehensive critique of methodological diversity, and a comprehensive examination of future research opportunities and standardization initiatives in this constantly changing field. This approach not only guarantees a transparent representation of the constraints of our investigation but also emphasizes the critical necessity of data standardization and enhanced collaboration within this specialized NER domain.

The literature search was limited to publications published between 2018 and 2024 that utilized NER-related keywords, including entity extraction, entity recognition, and NER, as well as the specified domain, specifically the government domain, in the title, abstract, and keywords. In line with PRISMA guidelines (Section: Method, Point 6—Information Sources), the results were obtained from the SCOPUS database using a systematic search strategy. It utilized predefined keywords and filters aligned with the research objectives. The database was accessed on 6 March 2024 to ensure the inclusion of the most recent and comprehensive studies (PRISMA, 2020). The search query is the following: TITLE-ABS-KEY ((“named entity recognition” OR “NER” OR “name entity recognition” OR “entity extraction”) AND “government”).

The initial search above produced 129 articles from reputable publishers, including ACM, Elsevier, IEEE, and MDPI. During the initial inquiry, several articles outside the domain of this study were identified. In certain articles, the term “NER” is also used to refer to the North-East Region rather than Named Entity Recognition. Hence, an additional article selection procedure was conducted.

Since we found some articles that were irrelevant to our survey study, we need to omit those several articles from our initial search results. In order to get the most relevant articles to our study, we applied three filtering steps as described in **Figure 1**.

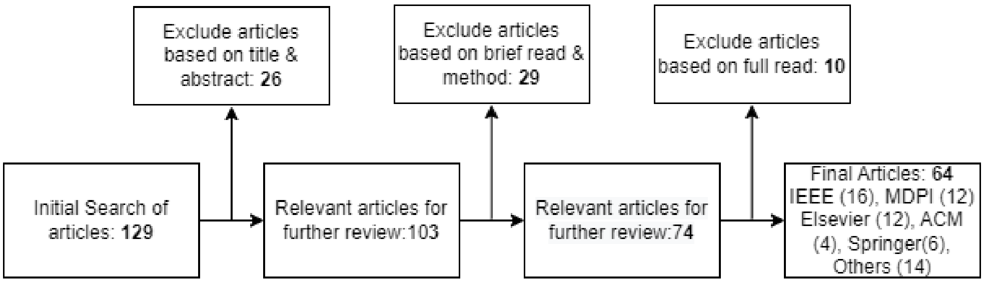

Figure 1. Articles selection process using three steps filtering.

The initial stage of the article selection process involves a meticulous review of both titles and abstracts extracted from the initial search results. To facilitate this task, we employed Mendeley software, a versatile tool for efficient article management. The software not only allows us to seamlessly import the query results but also aids in the systematic review of titles and abstracts, providing a vital preliminary insight into each article’s relevance. This first round of scrutiny led to the exclusion of 26 articles that proved irrelevant to our study objectives. For instance, articles pertaining to NER within the Northeast Region were rightfully omitted, given their divergence from our research focus. Consequently, we narrowed down our selection to 103 articles, which form the basis for our subsequent rigorous screening phase. In this subsequent stage, each of the remaining articles will undergo a thorough reading, allowing us to delve deeper into their contents, methodologies, and contributions to the NER discourse within the government domain.

After applying a brief read and method, we found 29 irrelevant articles to our study; for example, the word government in those irrelevant articles is not a domain of NER research but only as a funding organization for each article. Therefore, we omit those articles. From this process, there are 74 articles remaining to go through the next filtering process, which is a full read of each article.

In the domain of government-related NER, it is essential to clarify a fundamental assumption underpinning this study's article selection process. We operate under the premise that the scope of articles to be included in our analysis encompasses any research endeavor employing NER methods in the context of government institutions. This assumption recognizes the multifaceted applications of NER within the government domain, which may include but are not limited to information extraction from government documents, identification of public officials' names, or the extraction of administrative entities from textual data. By encompassing this broad spectrum of research applications, we aim to comprehensively evaluate the diverse landscape of NER techniques and their effectiveness in addressing the distinctive challenges posed by government-related text corpora. This assumption is a foundational principle in our study, guiding our systematic review process and shaping the inclusivity criteria for the articles under investigation.

We planned to analyze NER performances in the government domain by calculating the *F*-Score for each article as the harmonic mean between precision and recall, which is commonly used in NER research. After the full read of 74 remaining articles, we omit 10 articles in which we think the government or human resources domain is not involved or beneficially affected by the NER articles; therefore, we have 64 final articles collected for this study.

The strategic decision to employ the focused query TITLE-ABS-KEY ((“named entity recognition” OR “NER” OR “name entity recognition” OR “entity extraction”) AND “government”) was made to restrict search results to NER research in government-related topics. Expanding the query with additional terms such as “NLP” or “information retrieval” would have widened the scope, including articles that were either not centered on NER or not relevant to governmental use cases. Although this more restrictive approach restricts the inclusion of potentially pertinent studies discovered through the citation chain, it offers an initial dataset that is specifically tailored to the subject matter. SCOPUS was selected to ensure compatibility with bibliometric tools like Bibliometrix, thereby balancing automation and profundity in the review process.

A bibliometric analysis of the final selection of articles, which comprises 64 scholarly contributions, is a critical component of our analysis, in addition to the detailed synthesis method employed in this study. The utilization of bibliometric analysis was a valuable instrument for conducting statistical checks, which revealed significant insights into the landscape of Named Legal Entity Recognition (NER) research in the government domain. The analysis process entailed the construction of word clouds, the establishment of a co-occurrence network, and the monitoring of the annual increase in article counts. The library capabilities of the R tool, Bibliometrix, were employed to enable this comprehensive analysis. It is important to mention that the decision to utilize SCOPUS as our primary search source in this study was motivated by the necessity to comply with the structural requirements of Bibliometrix, which mandates data in *.bib format for further analysis. Notably, Bibliometrix is an open-source R library that was developed by Massimo Aria and Corrado Cuccurullo in 2017, as documented in their work (Aria and Cuccurullo, 2017). It is specifically designed for comprehensive scientific mapping analysis. Providing a valuable layer of insight, this tool substantially enhanced our capacity to dissect and comprehend the intricate network of NER research in the government domain.

3. Results and analysis

3.1. SLR results

After exploring the bibliometrics analysis results, we examined 64 articles as the SLR results. The articles were subsequently summarized and mapped according to the following criteria: author, task, governance field, data set, entity, method, and performance measurement. Following a thorough examination of the 64 articles in **Table 1**, we endeavored to categorize them into the various areas of governance, as illustrated in **Table 2**. The classification of government areas in this review is based on the specific institutions responsible for addressing the issues, as evidenced by the articles analyzed. This procedure was selected to correspond with the Indonesian authors' understanding of the structure of government operations. In Indonesia, the government's responsibilities are divided between central and regional agencies, each of which has distinct jurisdiction over specific areas. For instance, the legal department or the Ministry of Law and Human Rights typically administers public policy, while the House of Representatives (DPR) and the Office of National Unity and Politics, which is under the Ministry of Home Affairs, supervise political matters. Public policy and politics are classified separately in the analysis as a result of these distinct responsibilities. This classification guarantees that the investigation accurately represents the operational and administrative framework of government institutions.

Table 1. SLR results based on government field.

Government Field	Total	Articles
E-government	4	(Begen and Vidiasova, 2022), (Eva and Weber, 2022), (Ivanin et al., 2021), (Pena et al., 2018).
National security	10	(Barachi et al., 2022), (Cardenas et al., 2019), (Dominic et al., 2023), (Gasmi et al., 2019), (Gultiaev and Domashova., 2022), (Lutfi et al., 2021), (Park et al., 2020), (Sandescu et al., 2022), (Wang et al., 2023), (Zacharis et al., 2023).
Archive management	4	(Kim et al., 2022), (Maurel et al., 2019), (Xiong et al., 2021), (Zhu et al., 2022).
Public policy	7	(Al-laith and Shabaz., 2021), (Bönisch et al., 2023), (Heusden et al., 2023), (Liu et al., 2018), (Qi et al., 2022), (Sufi, 2022), (Zhang et al., 2021).
Human resources	2	(Ramdhani et al., 2021), (Toledo et al., 2020).
Public health	12	(Agarwal et al., 2022), (Azzouzi et al., 2022), (Bajaj et al., 2022), (Chandra et al., 2023) (Nemes and Kiss, 2021), (Shen and Spruit, 2021), (Silvestri et al., 2021), (Street et al., 2022), (Suthpin et al., 2022), (Tan et al., 2020), (White et al., 2022), (Wicaksono and Mariah, 2019).
Finance	5	(Dogra et al., 2021), (Jayakumar et al., 2020), (Kulkarni et al., 2023), (Pugliese et al., 2021), (Wang et al., 2022).
Tourism	1	(Bouabdallaoui et al., 2022)
Disaster management	2	(Han and Wang, 2019), (Nasution et al., 2022)
Procurement	4	(Guimarães et al., 2024), (Pimpisal et al., 2021), (Wang et al., 2018), (Yaozu and Jiangen, 2020).
Transportation	3	(Agarwal et al., 2019), (Li et al., 2023), (Suat-Rojas et al., 2022).
Agriculture	1	(Gangadharan et al., 2020)

Table 1. (*Continued*).

Government Field	Total	Articles
Politics	1	(He et al., 2019)
Law	6	(Bach et al., 2019), (Chen et al, 2020), (Garat and Wonsever, 2022), (Leitner et al, 2019), (Martinez-seis et al., 2022), (Niu and Zheng, 2018).
Environment	2	(Liu et al., 2023), (Mannavarasan et al., 2022).

As a result, we found that public health is the most dominant topic in government NER, as shown in **Figure 2**.

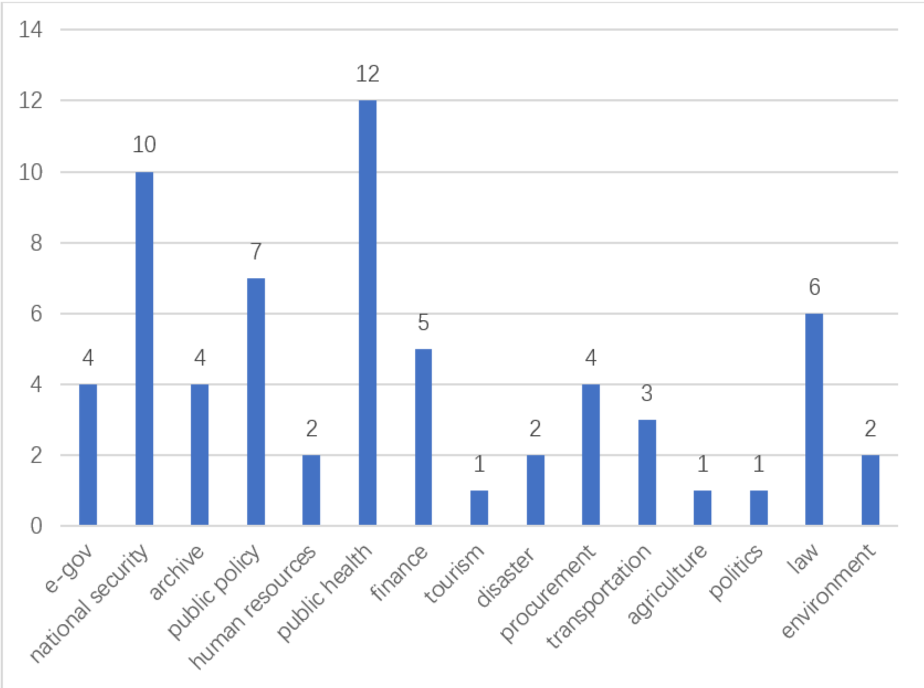

Figure 2. Number of articles for each field.

Table 2. SLR results based on proposed methods.

Method	Total	Articles
Rule-based	11	(Cardenas et al., 2019), (Chandra et al., 2023), (Gangadharan et al., 2020), (Han and Wang, 2019), (He et al., 2019), (Maurel et al., 2019), (Niu and Zheng, 2018), (Ramdhani et al., 2021), (Street et al., 2022), (Zhang et al., 2021), (Zhu et al., 2022),
Deep learning	19	(Agarwal et al., 2022); (Bajaj et al., 2022), (Begen and Vidiasova, 2022), (Bouabdallaoui et al., 2022), (Gasmi et al., 2019), (Ivanin et al., 2021), (Kim et al., 2022), (Li et al., 2023), (Liu et al., 2023), (Mannavarasan et al., 2022), (Pena et al., 2018), (Sandescu et al., 2022), (Shen and Spruit, 2021), (Silvestri et al., 2021), (Toledo et al., 2020), (Wang et al., 2022), (White et al., 2022), (Xiong et al., 2021), (Zacharis et al., 2023).
Machine learning	19	(Agarwal et al., 2019), (Al-laith and Shabaz., 2021), (Barachi et al., 2022), (Bönisch et al., 2023), (Dominic et al., 2023), (Garat and Wonsever, 2022), (Guimarães et al., 2024), (Heusden et al., 2023), (Jayakumar et al., 2020), (Lutfi et al., 2021), (Martinez-seis et al., 2022), (Nasution et al., 2022), (Nemes and Kiss, 2021), (Pimpisal et al., 2021), (Pugliese et al., 2021), (Suat-Rojas et al., 2022), (Sufi, 2022), (Wicaksono and Mariah, 2019), (Yaozu and Jiangen, 2020),
Combined	15	(Azzouzi et al., 2022), (Bach et al., 2019), (Chen et al, 2020), (Dogra et al., 2021), (Eva and Weber, 2022), (Gultiaev and Domashova., 2022), (Kulkarni et al., 2023), (Leitner et al, 2019), (Liu et al., 2018), (Tan et al., 2020), (Park et al., 2020), (Qi et al., 2022), (Suthpin et al., 2022), (Wang et al., 2018), (Wang et al., 2023)

Public health is the most frequently discussed topic on NER for the government field, with 12 articles. National security, public policy, and legal topics follow, with ten articles, seven articles, and six articles, respectively. Human resources and transportation produced only two articles, while tourism, agriculture, calamity, and politics each produced one article. The limited number of articles produced indicates that additional NER research is still required in the fields of human resources and administration. We also tried to classify the SLR results into four types of methods, as shown in **Table 2**.

As a result, machine learning and deep learning are the most popular methods used in government NER, as shown in **Figure 3**.

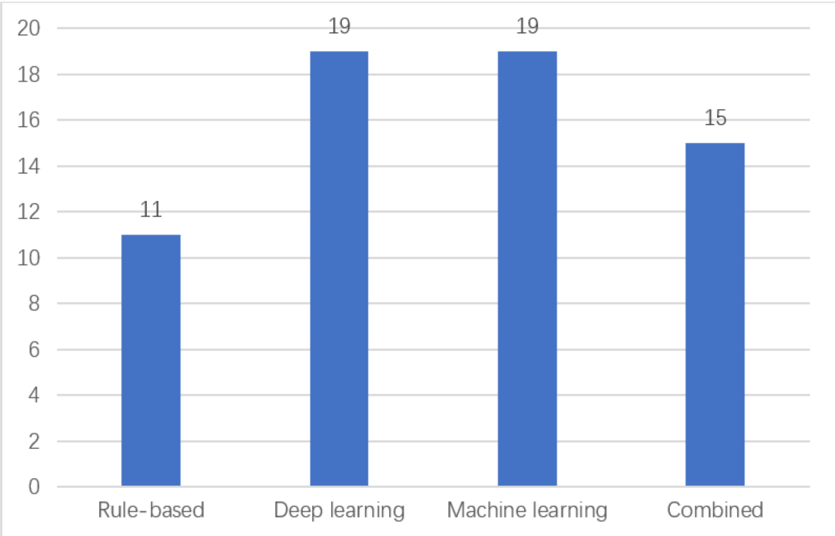

Figure 3. NER methods in the government domain.

Table 3. SLR result based on performance measurement.

Performance Measurement	Total	Articles
Accuracy, Precision, Recall, *F*-Score	45	(Agarwal et al., 2022), (Azzouzi et al., 2022), (Bach et al., 2019), (Bajaj et al., 2022), (Barachi et al., 2022), (Begen and Vidiasova, 2022), (Bouabdallaoui et al., 2022), (Chen et al, 2020), (Gangadharan et al., 2020), (Garat and Wonsever, 2022), (Gasmi et al., 2019), (Guimarães et al., 2024), (Gultiaev and Domashova., 2022), (Han and Wang, 2019), (Heusden et al., 2023), (Ivanin et al., 2021), (Jayakumar et al., 2020), (Kim et al., 2022), (Kulkarni et al., 2023), (Leitner et al, 2019), (Li et al., 2023), (Liu et al., 2018), (Liu et al., 2023), (Mannavarasan et al., 2022), (Martinez-seis et al., 2022), (Maurel et al., 2019), (Nasution et al., 2022), (Niu and Zheng, 2018), (Park et al., 2020), (Qi et al., 2022), (Ramdhani et al., 2021), (Sandescu et al., 2022), (Suat-Rojas et al., 2022), (Shen and Spruit, 2021), (Silvestri et al., 2021), (Sufi, 2022), (Suthpin et al., 2022), (Toledo et al., 2020), (Wang et al., 2023), (Xiong et al., 2021), (Zhang et al., 2021), (Zhu et al., 2022).
Error input rate	1	(Pugliese et al., 2021)
Not available	18	(Agarwal et al., 2019), (Al-laith and Shabaz., 2021). (Bönisch et al., 2023), (Cardenas et al., 2019), (Chandra et al., 2023), (Dogra et al., 2021), (Dominic et al., 2023), (Eva and Weber, 2022), (He et al., 2019), (Lutfi et al., 2021), (Nemes and Kiss, 2021), (Pena et al., 2018), (Street et al., 2022), (Tan et al., 2020), (Wang et al., 2022), (Wicaksono and Mariah, 2019), (Yaozu and Jiangen ,2020), (Zacharis et al., 2023).

Figure 3 shows that machine learning and deep learning methods are the most popular methods in NER for the government domain with 19 usages of the model, followed by combined methods with 19 usages and rule-based method with 11 usages. Using machine learning and deep learning models on NER for the government domain is preferable, considering a good performance measurement result from machine

learning models such as CRF or SVM. Ten articles used combined methods of deep learning and machine learning, such as BiLSTM + CRF (Liu et al., 2018) and LSTM + ILP (Wang et al., 2018), or deep learning and rule-based such as Distilbert + Rule-based (Dogra et al., 2021). The most popular combined method in NER for government is BiLSTM + CRF, with 6 of 10 articles. In terms of performance measurement, we found that not every article provides a performance measurement, as shown in **Table 3**.

We found that performance measurements were not provided in 28.13% (18 articles). When we examine these 18 articles in greater detail, we observe a trend that the utilization of tools like Spacy or Stanford NLP does not involve performance measurement. A diverse array of performance metrics have been implemented. Accuracy, precision, recall, and *F*-score are the most frequently employed metrics. However, the input error rate was employed as a performance metric in one article. The instrument Microsoft Power Business Intelligence was employed to extract entities associated with the source of negative news, resulting in the highest *F*-Score of 99.4% (Sufi, 2022). Using the BERT deep learning model, the lowest *F*-Score was achieved at 57% (Ivanin et al., 2021). Although the tool achieved the highest measurement and the deep learning model achieved the lowest *F*-Score, this does not imply that the method is the sole factor contributing to the performance measurement results. Some deep learning models can also generate high-performance measurements, such as the BERT and RoBERTa models, which can generate an *F*-score of 93% to extract drug-related entities (Bajaj et al., 2022) or the Elmo and BERT models, which can generate an *F*-score of 96.61% to extract biomedical entities in medical records in Italy (Silvestri et al., 2021).

On the contrary, the use of NER tools in the government domain does not guarantee the most accurate performance measurement results. For instance, the use of Spacy and CoreNLP tools only generates an 89 93% *F*-Score on NER legal documents (Garat and Wonsever, 2022). Furthermore, the use of NER tools is frequently incapable of generating performance measurements, as evidenced by the use of Spacy tools on Covid-related tweets (Nemes and Kiss, 2021) and the use of NER to extract entities related to Indonesian public health from 1204 news from 2015 to 2019 (Wicaksono et al., 2019). Additionally, we identified eight articles that implemented the combined BiLSTM + CRF model, resulting in measurement outcomes that ranged from 95.95% to 81.7%. Based on the performance measurements of these eight articles, this combined method should be taken into account for application in the human resources and government sectors. Some datasets and tools or source codes are publicly available, as shown in **Table 4** and **Table 5**.

Table 4. Available source code or NER tools.

Articles	Source Code or Tools	URL
(Al-Laith and Shahbaz, 2021)	Farasa Arabic NER tool	http://qatsdemo.cloudapp.net/farasa
(Bönisch et al., 2023)	Bundestag-Mine	https://bundestag-mine.de/
(Guimarães et al., 2024)	DODFMiner	https://www.dodf.df.gov.br/
(Zhu and Cole, 2022)	PDFDataExtractor	https://github.com/cat-lemonade/PDFDataExtractor

Table 4. (*Continued*).

Articles	Source Code or Tools	URL
(Suat-Rojas et al., 2022)	Spacy Library	https://explosion.ai/
(Maurel et al., 2019)	Unitex NER Tool	https://unitexgramlab.org
(Shen and Spruit, 2021)	Source Code	https://github.com/ianshan0915/ade-extraction
(Bach et al., 2019)	Source Code	https://github.com/trungtv/pyvi
(Han and Wang, 2019)	Source Code	http://ictclas.nlpir.org/

Only 9 articles provided tools or source code publicly with 5 articles using NER tools such as Farasa, Spacy Library, Unitex, Bundestag-Mine, and DODFMiner, while the remaining 4 articles wrote their own source code and uploaded it on Github or their own websites as shown in **Table 4**.

In the context of a systematic literature review for the NER research government domain, it is significant to note that only a small number of 12 scholarly articles generously share their accompanying datasets with the public, as illustrated in **Table 5**. This illustrates the significance of data accessibility and transparency in NER research, as it facilitates the field's advancement, validation, and reproducibility. These pioneering researchers are establishing a valuable precedent for the broader academic community, promoting collaboration and facilitating the development of more accurate and efficient NER models and techniques.

Table 5. Available dataset.

Articles	Dataset	URL
(Bajaj et al., 2022)	Drug trends report	https://mha.ohio.gov/Researchers-and-Media/Workgroups-and-Networks/Ohio-Substance-Abuse-Monitoring-Network
(White et al., 2022)	Hmong Medical dataset	https://www.health.state.mn.us/communities/translation/hmong.html
(Al-Laith and Shahbaz, 2021)	Verified twitter accounts	https://www.dropbox.com/s/9pxhr8k8nthdh7z/TwitterAccounts.txt
(Sandescu et al., 2022)	Cybersecurity news platform	https://yggdrasil.codaintelligence.com/dataset.xlsx
(Ivanin et al., 2021)	Russian government documents	https://github.com/dialogue-evaluation/RuREBus/ dataset
(Agarwal et al., 2019)	Indian locations	https://en.wikipedia.org/wiki/Category:Roads_in_Delhi, www.flooraddress.in
(Pimpisal et al, 2021)	Thai procurement documents	https://opendata.nesdc.go.th/dataset/emenscr-project
(Yaozu and Jiangen, 2020)	Chinese procurement documents	http://www.ccgphebei.gov.cn/xt/xt_wx/cggg/zhbggAAAA/201909/t20190925_1114497.html
(Shen and Spruit, 2021)	Medicines documents	https://www.medicines.org.uk/emc
(Leitner et al., 2019)	Court documents	https://github.com/elenanereiss/Legal-Entity-Recognition
(Wang et al., 2018)	Chinese court documents	http://wenshu.court.gov.cn/
(Martinez-seis et al., 2022)	Mexical court documents	https://cloud.upiita.ipn.mx/docentes/s/EgY4sQNmZNQ9iNx/authenticate/showShare

3.2. Bibliometric analysis results

The *.bib file format is utilized by both the Mendeley software and the Bibliometrix R library. However, they are incompatible due to their distinct structures. Consequently, we exclusively employed the Mendeley *.bib format to peruse abstracts and titles during the initial phase of the article selection process. We employed the original *.bib format of the SCOPUS query results for additional bibliometric analysis, as it contains additional information, including the number of citations, article type, references, affiliations, and funding information.

One of our study objectives is to gain insight into NER research in the government domain, and we exclusively employ a number of features of the Bibliometrix library to achieve this. The annual growth diagram, co-occurrence network, word cloud, and region map are the features that indicate which country has the highest number of articles related to NER research in the government domain. The annual growth diagram shows a time series of article numbers during the 7 years from 2018 to 2024, as shown in **Figure 4**.

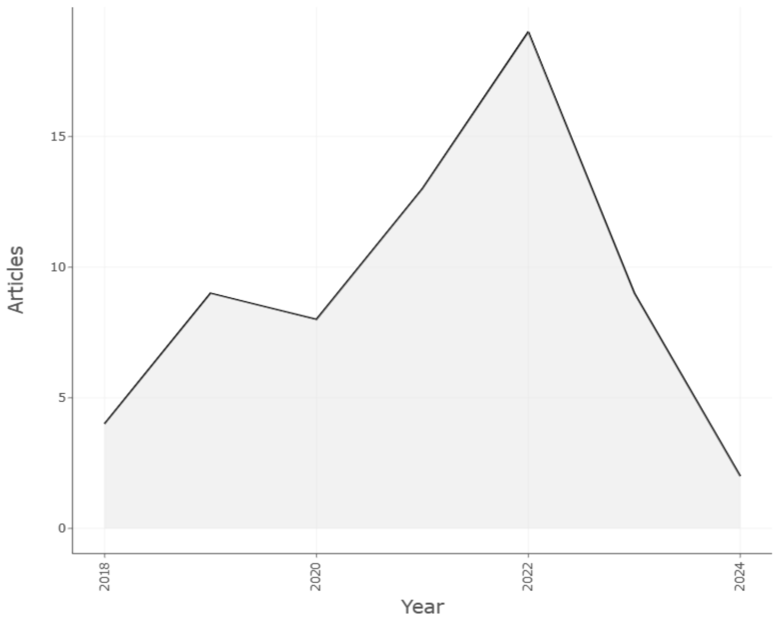

Figure 4. The annual growth of NER governmental research.

NER government research has a positive trend, as evidenced by the annual growth diagram in **Figure 4**. The number of articles has increased from 4 in 2018 to 13 in 2021, and it has reached its highest point of 18 in 2022. However, the number of articles has slightly decreased in 2023, with only 9 in 2023 and only 2 in early March 2024. From this diagram, it is evident that the number of articles on NER government research topics increases over the course of five years (2018–2022) and decreases in 2023 (9 articles) and early March 2024 (2 articles). The co-occurrence network was employed to offer insights into the structure, themes, and collaborations of NER government research topics, as illustrated in **Figure 5**.

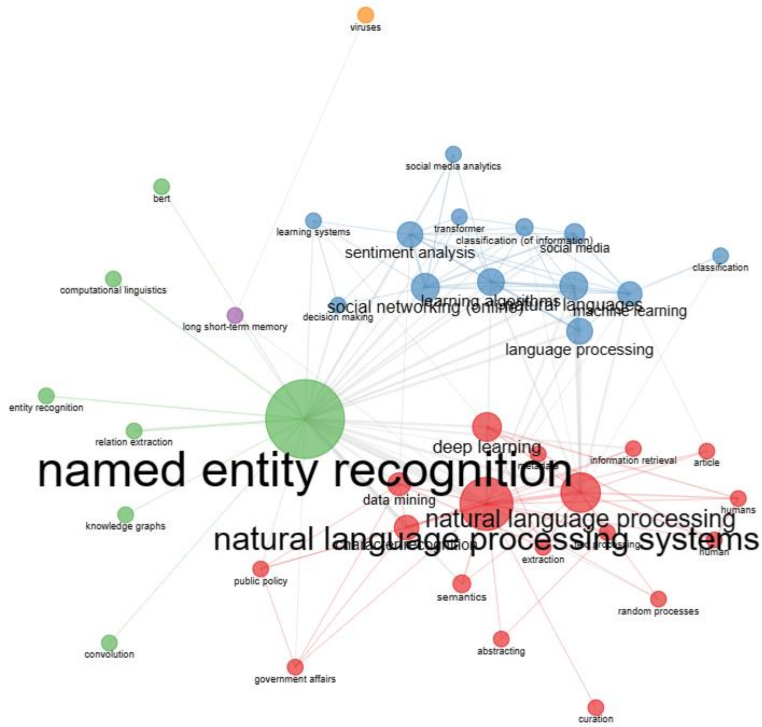

Figure 5. The co-occurrence network of NER governmental research.

Based on **Figure 5**, the co-occurrence network of government NER research is dominated by five topic clusters: the named entity recognition cluster (represented by green dots), the natural language processing system cluster (represented by red dots), the online social network cluster (represented by blue dots, purple dots, and orange dots), and viruses (represented by orange dots). In the character recognition cluster, topics related to government are present in multiple clusters, including government affairs and clever government affairs services. They also appear as decision makers in the online social network cluster and as e-government in the online named entity recognition cluster. The word cloud feature was also employed to obtain an understanding of the terms that are frequently referenced in NER government topics, as illustrated in **Figure 6**.

Figure 6. The word cloud of NER governmental research.

Figure 6 shows that several keywords are occasionally used in this study, such as natural language processing, online social networking, data mining, and character recognition. The size of the words shows how often the keywords appear within the NER governmental research. Government affairs and public policy are among the government keywords that appear infrequently in this word cloud, as indicated by their word size. We can conclude that NER for the government domain is still uncommon, as shown in **Figures 5** and **6**. Consequently, additional research is required in the government domain regarding NER.

The final feature that we employ from the Bibliometrix library for this study is the Country Scientific Production feature, which indicates the country in which NER governmental research is frequently conducted, as illustrated in **Figure 7**.

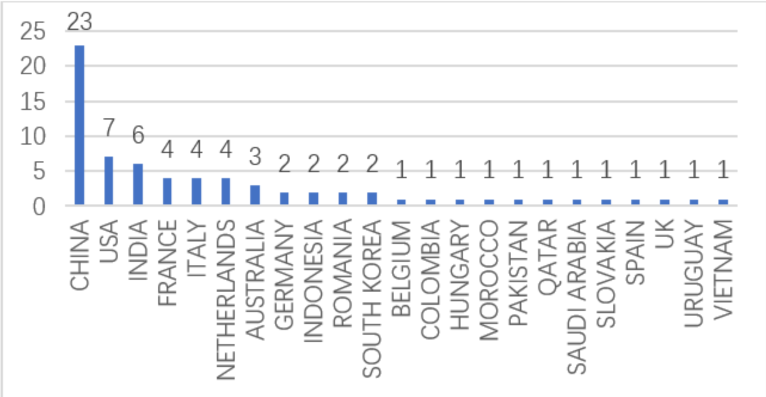

Figure 7. The country's scientific production of governmental NER.

3.3. Case studies of NER application in government

Several studies have demonstrated the substantial influence of NER applications on the enhancement of government operations. For instance, the human resources department of a local government in Indonesia has simplified document management by automating the extraction of important entities for the HR database through the implementation of a rule-based NER system. In comparison to the manual process, which used to require over a minute, the system has reduced the extraction time to 0.044 seconds per document (Ramdhani et al., 2021). The NER tool specifically designed for Italy was able to reduce invoice entry errors from 10% to 2% and increase the daily capacity of invoice processing from 30 to 100 (Pugliese et al., 2021). In addition, the Bogota city government implemented NER to enhance traffic accident detection, extract location data from social media, and augment existing methods with real-time Twitter reports (Suat-Rojas et al., 2022). These examples demonstrate the substantial enhancements in operational accuracy and service efficiency that can be achieved through the implementation of NER in government.

4. Discussion

The subject of public health in government NER appears to have been the most frequently researched over the past five years, as evidenced by the publication of twelve articles. One explanation is that the COVID-19 virus's global pandemic at the end of 2019 has resulted in a significant amount of COVID-19-related research, including NER research for the government domain. This research aims to extract public health-related entities from a variety of sources, including news, social media, and government reports. In contrast, there still needs to be more NER on the subject of human resources in the government sector. Therefore, additional research is still necessary in this field to address the remaining unresolved issues. For instance, the diversity of language employed in job postings and resumes presents a challenge in identifying specific entities pertinent to the human resources domain, such as job titles, talents, and certifications. Furthermore, accurate entity recognition necessitates more advanced natural language understanding techniques, as it is essential to comprehend the context in which these entities are employed. In general, additional research is required to create NER models that are particularly designed for the human resources domain. This will enhance the recruitment and talent management process, as well as the human resources document management process.

In the government sector, deep learning has become a widely accepted and sophisticated method of NER. NER tasks, in which the language is highly variable and context-dependent, notably benefited from the capacity of deep learning models to learn complex patterns and relationships in large data sets. This is the primary reason for this. In a variety of domains, including the government human resources domain, deep learning models such as Transformer models (BERT, GPT) have been demonstrated to attain state-of-the-art performance on NER tasks. Additionally, these models can be trained on extensive quantities of labeled data, which are frequently accessible in human resource and government datasets.

Furthermore, deep learning models can be tailored to perform particular tasks, enabling them to acquire the ability to identify named entities that are unique to the human resources domain, including job titles, skills, and qualifications. In general, deep learning models are a promising approach for NER in the government human resources domain because of their capacity to learn complex patterns in large datasets and their ability to fine-tune specific tasks. The most frequently employed approaches in NER for government are still deep learning and machine learning, with 19 articles. This is followed by combined methods, which have 15 articles, such as BiLSTM + CRF and LSTM + ILP methods.

One of the most effective methods to measure NER performance is calculating the *F*-Score, the harmonic mean of precision and recall. Unexpectedly, the highest *F*-Score from this study was produced using Microsoft Power Business Intelligence tools, with 99.4% (Sufi, 2022). This seems aberrant since more than 50% (7 of 13 articles) of the other research used tools do not provide any performance measure. We found it aberrant since the same tool, such as the Spacy library, actually produces *F*-Score in some of the articles (Garat et al., 2022; Suat-Rojas et al., 2022), yet other research which also uses the Spacy library does not provide any performance measurement (Nemes and Kiss, 2021; Wicaksono and Mariah, 2019). Even though deep learning has been the popular method for NER in the government domain (17 articles), this study's lowest measurement was produced using this method with only 57% *F*-score (Ivanin et al., 2021).

We tried to calculate the performance results for each method by not including research with no performance results and the result without the *F*-score measurements (error input rate), as shown in **Figure 8**.

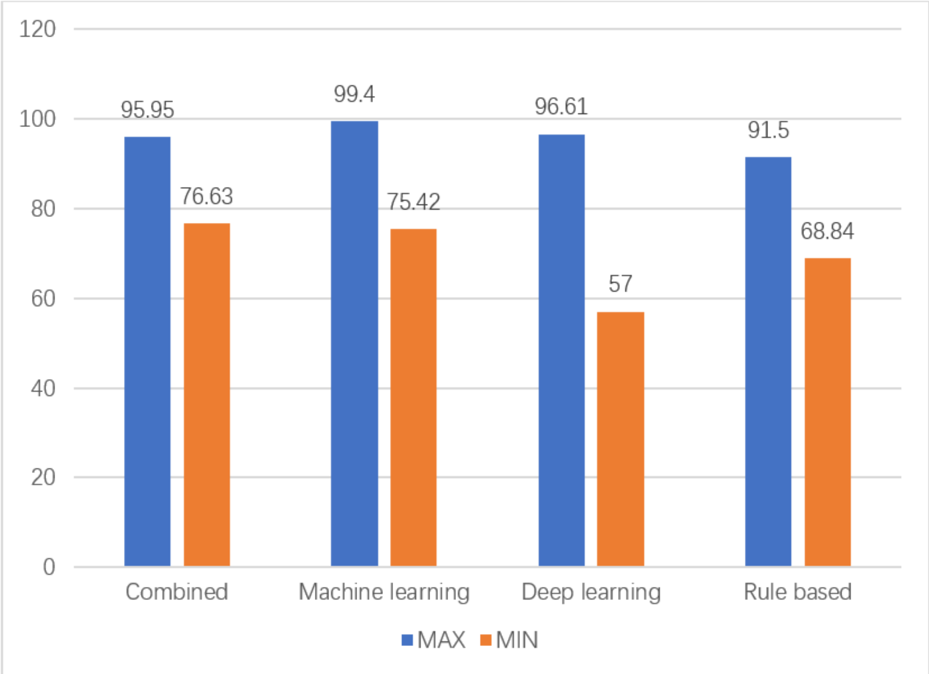

Figure 8. NER methods performance comparison.

The highest performance measurement result using machine learning is 99.4%, and the lowest result is 75.42% from 50% of articles using machine learning for NER in the government domain. A total of seven articles were excluded due to the absence of performance measurement results obtained through machine learning techniques. The combined procedure produced the highest performance measurement result of 95.95% and the lowest result of 76.63% for 15 articles. From the 11 articles, the rule-based method yielded the highest performance measurement result of 91.5% and the lowest, which was 68.84%. Of the 19 articles, the performance measurement result article using the deep learning method for NER in the government domain was the highest at 96.61% and the lowest at 57%. Only one article did not include the performance measurement result article. We employed a stringent methodology to assess the performance of a variety of methods, with a particular emphasis on the utilization of maximum and minimum *F*-score values. The deliberate selection of this methodology was motivated by the inherent variability of the datasets employed in the various articles. Averaging *F*-scores across various datasets can result in potentially misleading conclusions, as it is acknowledged that datasets exhibit a wide range of complexity, size, and domain. To provide a more accurate representation of their capabilities when confronted with various datasets, we assure a more robust evaluation of their performance by taking into account the extreme *F*-score distributions for each method. This methodological decision enhances the reliability and validity of our comparative analysis, enabling users to make more informed decisions when selecting appropriate techniques in the context of diverse data-driven research.

Tools such as Microsoft Power Business Intelligence, Stanford NLP, Farasa, D2RQ, Unitex, BERN, CoreNLP, Freeling, IBM Watson, and Spacy library are used to generate certain methods, including machine learning and combined methods between machine learning and deep learning. Six of the thirteen articles utilized the Spacy library, which is predicated on the CRF machine learning method, due to the fact that these instruments are not regarded as methods in NER. We endeavored to meticulously examine the methodologies employed by each instrument in order to determine the appropriate method for comparing the performance of the measurements, as previously described.

The performance comparison of the methods in **Figure 8** appears to be unfair, as each method employs distinct datasets and a variety of entities. Therefore, we attempted to compare the performance of the methods that employ common entities, as defined in the 6th Message Understanding Conference (MUC) when NER was first introduced, including persons (PER), locations (LOC), organizations (ORG), date (DAT), and time (TIME). (Grishman and Sundheim, 1995). We only compare the performance of this method on results that include detailed measurements for each entity; as a result, certain results are excluded. The performance comparison method using F1-Score measurements for each similar entity is shown in **Table 6**.

Table 6. Similar entities method performances comparison.

Articles	Methods	PER	ORG	LOC	DAT	TIME
(Pimpisal et al., 2021)	Machine Learning	-	93.68%	-	-	-
(Garat and Wonsever, 2022)	Machine Learning	90.21%	-	-	-	-
(Suat-Rojas et al., 2022)	Machine Learning	-	-	91.77%	-	68.37
(Sufi, 2022)	Machine Learning	-	-	99.59%	-	-
(Wang et al., 2018)	Deep Learning	96.89%	86.09%	52.9%	-	-
(Gasmi et al., 2019)	Deep Learning	-	93%	-	-	-
(Toledo et al., 2020)	Deep Learning	83%	68%	50%	-	-
(Xiong et al., 2021)	Deep Learning	95.6%	88.05%	84.07%	87.11%	-
(Silvestri et al., 2022)	Deep Learning	-	98.69%	-	97.96%	-
(Bajaj et al., 2022)	Deep Learning	93%	88%	73%	98%	78%
(Li et al., 2023)	Deep Learning	-	-	94.59%	-	-
(Leitner et al., 2019)	Combined	95.41%	79.89%	-	-	-
(Chen et al., 2020)	Combined	83.68%	70.78%	80.6%	-	-
(Azzouzi et al., 2020)	Combined	97.36%	-	-	98.32%	-

Table 6 highlights the best F1-Scores achieved for different entity types, indicating their highest performance. In particular, the combined procedure produces superior F1 scores for extracting entities such as PER (97.36%) and DAT (98.32%). The most favorable results are achieved by deep learning algorithms for ORG (98.69%) and TIME (78%). In the interim, LOC (99.59%) obtains its highest F1-Score by employing machine learning techniques. The efficacy of hybrid combined methods is notably evident in the extraction of commonly used entities such as PER and DAT.

Table 6 outlines the common entities frequently extracted from government documents; however, this also highlights a limitation in applying NER within this domain. The extracted entities can vary significantly based on the specific government field under consideration. Notably, not all studies address or provide a consistent set of common entities, underscoring the variability and field-specific nature of NER implementation in government contexts.

We also examined which articles proposed new methods that could serve as inspiration for the direction of future development of new methods, particularly for NER methods. We identified at least three articles that endeavor to create new methods in the government domain by combining existing methods or constructing new rule-based algorithms. The Agriculture Named Entity Recognition using Topic Modeling techniques (AERTM) algorithm is a specialized algorithm that utilizes the Latent Dirichlet Allocation (LDA) algorithm to extract agricultural entities (disease, fertilizer, soil, pathogen) from 3000 sentences on various government websites. Without any comparison to other NER methods, this algorithm achieves an accuracy of 80%. A new method, BiLSTM + CRF + CT, is a combination of three distinct methods. In the initial layer, the deep learning method BiLSTM is combined with CRF in the subsequent layer. The output of the CRF layer will be trained on sentences that contain tags using compound labeling. The performance measurement produced by this technique is higher (76.63%) than that of the BiLSTM + CRF method (74.99%) and the CRF method (73.38%). GovAlbert, an additional novel combined method, integrates CRF with Albert's NER model. This technique yielded the highest performance measurement, with an *F*-Score of 89.4%, in contrast to Albert's 88.49%, Albert + BiLSTM + CRF's 88.92%, and Albert + BiLSTM's 87.52%. These methods demonstrate that the proposed NER method can be developed by combining existing methods, such as deep learning and machine learning, or by creating novel rule-based algorithms.

Common knowledge is that English is a well-known resource language for NLP, particularly in the field of NER. English is one of the most extensively researched and well-equipped languages for NER. English has garnered substantial attention and resources, rendering it an ideal candidate for NER development (Han et al., 2018). Conversely, Bahasa Indonesia and other languages are regarded as under-resourced in NER research. The scarcity of resources, such as annotated corpus, gazetteer, and computational tools, is to blame for the underdevelopment of Named Entity Recognition (NER) for Indonesian.

Furthermore, the intricacy of the language and the wide range of named entities it encompasses further contribute to this issue. Therefore, we examined how the language of the dataset in this study affects the performance results. The use of languages in the dataset is shown in **Figure 9**.

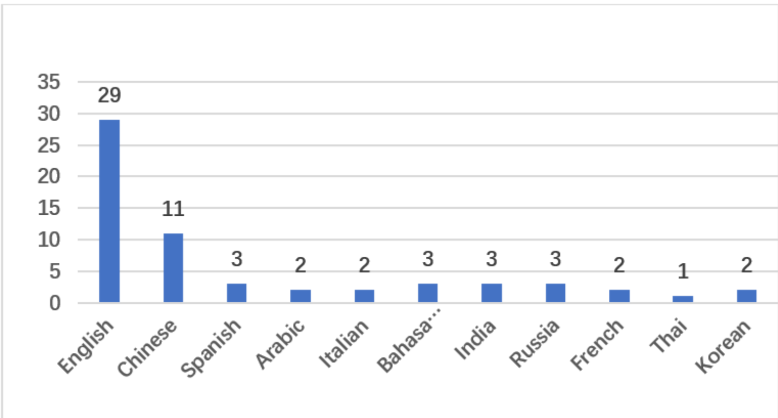

Figure 9. Dataset languages used in articles.

As previously mentioned and illustrated in **Figure 9**, English is also a frequently used resource language for NER in government. This is demonstrated by the 29 articles that utilized English in their dataset, followed by 11 articles in Chinese and three articles in Spanish, Indian, Russian, and Indonesian. The remaining languages (Arabic, Italian, French, Thai, and Korean) are classified as low-resource languages, as they contain fewer than three articles per language, particularly in the government domain of NER. English is a well-known resource language in NER, with an optimum performance of 99.4% and a minimum performance of 67% in this study. The optimum performance of Chinese was 95.4%, while the minimum performance was 76.63%. Spanish achieved an optimum performance of 98.04% and a minimum performance of 96.85%. 57% of the articles were generated in Russian, which resulted in the lowest average performance. The low-performance rate (57%) does not substantiate the claim that low-resource languages, such as Russian, also yield low NER performance. For example, Italian, which was only used once in this study, was able to generate an *F*-score of 96.61%. In NER for government domains, low-resource languages can achieve superior performance compared to well-known resource languages like English or Chinese, as evidenced by the average performance figures by language in the dataset. We discovered in this investigation that a resource language that is widely recognized only sometimes results in superior performance when contrasted with a low-resource language.

We attempted to identify specific entities for NER in the government domain by excluding entities that are frequently used in the general domain, such as person, location, and organization. This was done in order to evaluate the impact of specific entities on performance measurement results in the government domain for NER. We identified nine articles that attempted to extract only person, location, and organization entities from six articles. Three of these articles did not include any performance measures. The generic entities' respective performance levels were as follows: 94.59% for the highest and 80.15% for the lowest. In addition, we discovered that 18 of the 35 articles with specific entities included person, location, and organization entities for extraction. As a result, only 17 articles extracted specific entities without person, location, and organization. Twelve of the 35 articles that contained specific entities lacked performance metrics. The minimum *F*-Score was 57%, while the maximum *F*-Score was 99.4% for the remaining 23 articles that contained specific entities.

The necessity of refining the model is an additional substantial constraint in the application of NER in the government sector, in addition to the challenges previously mentioned. Pre-trained models frequently only achieve optimal performance with additional training due to the specificity of entities in this domain, such as unique administrative terms or policy-related phrases. Adapting these models to recognize domain-specific entities effectively necessitates refinement. This process necessitates not only domain-specific annotated data but also the ability to configure and train the models, which can be resource-intensive. Additionally, refinement may elicit apprehensions regarding budgetary constraints, as it necessitates the provision of computing resources and expert labor to establish and sustain high-quality annotated datasets. Additionally, the application of NER in this sector may be further complicated by the potential emergence of data privacy and security concerns when managing sensitive government data for training purposes.

Custom-trained models, such as fine-tuned models, typically outperform their general counterparts in specific tasks. For instance, a study conducted in China demonstrated that the pre-trained Albert model attained an F1 score of 88.2%, while the fine-tuned version, Albert + 34,000, enhanced performance to 89.4%, indicating a 1–2% improvement (Xiong et al., 2021). The fine-tuned BERT model achieved an F1 score of 80%, surpassing common models such as FROG and Polyglot, which achieved lesser scores of 73% and 46%, respectively (Toledo et al., 2020). This trend is consistent with other models. This comparison underscores the fact that the performance of pre-trained models is enhanced when they are fine-tuned for specific datasets and tasks.

The integration of bibliometrics as a synthesis method within the framework of a Systematic Literature Review (SLR) has been demonstrated to be an invaluable method for enhancing our research findings on NER in government. We have been able to extract significant insights into the growth trajectory of NER research in this area through bibliometric analysis, which has highlighted the evolving landscape of scholarly contributions and annual trends. Additionally, the development of co-occurrence networks has uncovered a complex network of interrelated concepts and themes in NER discourse, thereby illuminating critical research clusters. Additionally, the word cloud diagram has effectively represented the critical keywords and themes that are prevalent in this field.

Lastly, the integration of country-level scientific production maps has provided a geographical perspective on the distribution of research activities, thereby facilitating a comprehensive comprehension of the global footprint of NER research in government domains. These bibliometric tools have complemented our SLR results, enhancing the depth and breadth of our analysis and fostering a more comprehensive comprehension of the field’s current state and evolution. Importantly, our synthesis method aligns with the PRISMA 2020 checklist, setting our approach apart from other SLR frameworks.

The findings from these case studies can be broadly applied across various government institutions to enhance operational efficiency. For instance, the human resources department can considerably reduce the time required for document management by utilizing NER for automated data extraction. Similarly, the finance department can optimize invoice processing and reduce errors by incorporating NER. Furthermore, the government’s capacity to promptly identify traffic incidents can be enhanced through the implementation of NER to extract location data from social media. These examples underscore the significant opportunity for government agencies to enhance service delivery in terms of accuracy and time efficiency by implementing NER technologies.

Named Entity Recognition (NER) has been demonstrated to be a valuable instrument in the automation and optimization of a variety of administrative processes within the context of government agencies. Nevertheless, the successful implementation of NER necessitates a tailored strategy that considers the distinctive attributes of various government sectors. The preliminary Government NER Ontology design has been created to aid government practitioners and researchers in the application of NER to specific domains.

The ontology demonstrates that Government NER can be tailored to extract both common entities (person, location, organization, date, time) and domain-specific entities, depending on the characteristics of the field. By utilizing suitable NER Methods and leveraging relevant Datasets, the system ensures optimal performance, which is measured through predefined metrics. Additionally, datasets are tailored to align with the specific needs of government fields, ensuring accurate and context-relevant entity extraction. This ontology provides a foundational framework that can be adapted and expanded by researchers and practitioners to enhance the application of NER in various governmental domains, as illustrated in **Figure 10**.

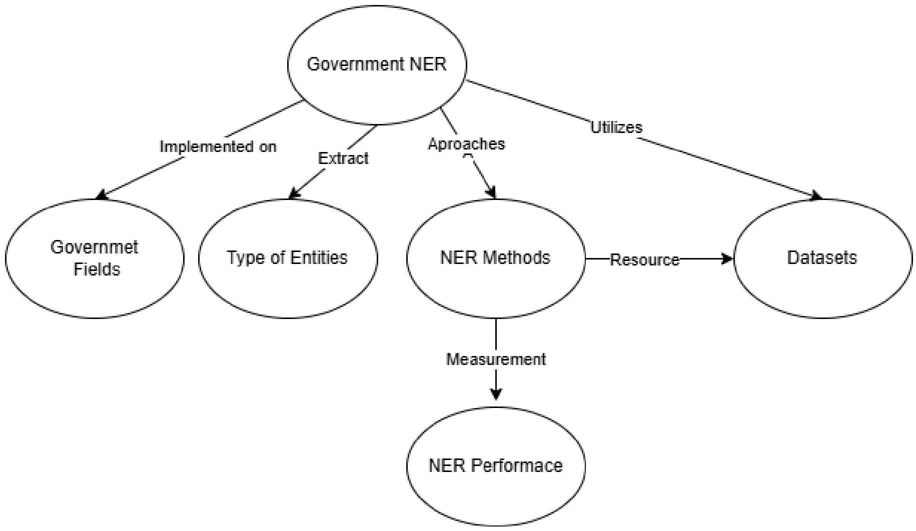

Figure 10. Preliminary government NER ontology design.

Government NER is designed to be applied across various Government Fields, each of which has unique characteristics and requirements. These fields—such as healthcare, public administration, finance, and law enforcement—can benefit from NER by optimizing the extraction and use of relevant entities within their specific contexts.

The primary objective of Government NER is to extract entities (Entity Types). These entities can be common entities, such as person, location, and organization, which are frequently encountered across different domains. Additionally, it can extract more specific entities, depending on the particular government field in which it is applied. For instance, in the healthcare field, specific entities might include patient identifiers, medical conditions, or drug names.

To perform this extraction, NER Methods serve as the approaches or techniques used to identify and extract entities. The effectiveness of these methods is evaluated through NER Performance, which is the measurement outcome of how well the methods perform in extracting the relevant entities.

Furthermore, Datasets are essential resources required by NER methods for training and testing the entity extraction process. These datasets are not only utilized by the NER system but are also tailored to align with the characteristics of specific Government Fields, ensuring the extracted entities are accurate and contextually relevant to the administrative needs.

5. Conclusion

5.1. Summary of findings

Public health was the most frequently researched topic in government, with 12 articles. National security and law were the second most researched topics, with ten articles. Public policy was the third most researched topic, with seven articles. Law was the fourth most researched topic, with three articles. Human resources, e-government, archives, tourism, disaster recovery, transportation, agriculture, politics, and the environment were among the topics that had fewer than three articles. The diverse applications of NER in government and the optimization of a variety of government activities are demonstrated by these 15 areas of NER.

The highest performance in NER in the government domain in this study reached as high as 99.4% using a tool called Microsoft Power Business Intelligence to extract entities related to the cause of negative news. Nevertheless, performance measurement was not feasible in seven of the studies that employed the instrument in this investigation. We discovered that the combined procedure had the highest quality for extracting PER (97.36%) and DAT (98.32%) when we compared the extracted entities to the first five introduced entities. Furthermore, we discovered that the deep learning method is the most effective in extracting ORG (98.69%) and TIME (78%), while the machine learning method is the most effective in extracting LOC (99.59%). These insights emphasize the effectiveness of extraction techniques that are specifically designed for various categories of similar entities in a government context.

As the most popular (19 usages) and state-of-the-art methods, deep learning or combined deep learning NER models should be taken into consideration, given the good performance measurement of SLR results. However, comparing other methods such as rule-based, machine learning, or NER tools with deep learning methods would be fair in proving which method is better for using the same dataset in the government domain.

The NER in Government domain research increased over the years (2018–2024), especially in China with 23 articles, followed by the United States with seven articles, India with six articles, and France, Italy, and the Netherlands with three articles. These results were obtained using a bibliometric analysis tool called Bibliometrix.

Based on the limited number of studies conducted, the NER for government contains numerous unresolved issues, including records management, transportation, and human resources, which each has two articles, and tourism, disaster management, politics, and agriculture, which each has one article.

The developed preliminary Government NER ontology design is intended to assist governmental institutions in implementing Named Entity Recognition (NER) that is tailored to specific administrative fields. Furthermore, this preliminary design can be leveraged and maintained by government practitioners and researchers to support current and future projects, fostering adaptability and continuous improvement in NER applications across various government domains.

5.2. Future directions

The efficacy of previous research can be enhanced by utilizing deep learning models such as BERT or combined deep learning models such as BiLSTM + CRF. In comparison to rule-based methods, deep learning models, such as BERT, and combined models, such as BiLSTM + CRF, have demonstrated promising results in enhancing the performance of Named Entity Recognition (NER) tasks. This is due to the fact that these models are capable of recognizing intricate patterns and relationships in the data, which may be challenging to document using conventional rule-based methods. Furthermore, the efficacy of deep learning models can be enhanced by their ability to customize them for specific tasks and to utilize large quantities of data. These models have demonstrated exceptional performance on NER tasks and have become a popular choice among NLP practitioners.

The limited scope of the SCOPUS database in this study presents an opportunity for future researchers to expand the dataset by incorporating articles from other databases. Broadening the range of sources will enhance the comprehensiveness of the literature, leading to more robust and reliable findings in the systematic literature review (SLR).

Incorporating expert validation for government field classification in future research will significantly enhance the robustness of this study. By consulting subject matter experts from diverse administrative and organizational contexts, the classification framework can be refined to account for variations across different countries and governmental structures. This approach will ensure that the study's findings are not limited by the specific characteristics of a single nation's organizational framework, thereby increasing its generalizability and applicability to a broader range of contexts.

Funding: This study was supported by research grants from Universitas Indonesia (Hibah Riset Internal Fakultas Ilmu Komputer Tahun Anggaran 2024-2025 Nomor NKB-6/UN2-F11.D/HKP.05.00/2024) and the Bogor Local Government of Indonesia (Kegiatan Pengembangan Kompetensi ASN Tahun Anggaran 2024 Sub Kegiatan Pengelolaan Lanjutan ASN No 5.03.02.2.03.04 Kode Rekening 5.1.02.02.11.003 Jurnal Internasional).

Conflict of interest: The authors declare no conflict of interest.

References

Agarwal, A., Toshniwal, D. (2019). Face off: Travel Habits, Road Conditions, and Traffic City Characteristics Bared Using Twitter. IEEE Access. https://ieeexplore.ieee.org/stamp/stamp.jsp? arnumber=8715356

Agarwal, I.Y, Rana, D.P, Shaikh,M, Poudel,S. (2020). "Spatio-temporal approach for classification of COVID-19 pandemic fake news," Social Network Analysis and Mining, Springer Vol 12 No 1, 2022Jayakumar, H, Krishnakumar, M.S, Peddagopu, V.V.V, Sridhar, R. RNN based question answer generation and ranking for financial documents using financial NER. Sadhana - Academy Proceedings in Engineering Sciences. Springer. https://link.springer.com/article/10.1007/s12046-020-01501-3

Al-Laith, A, Shahbaz, M. (2021). Tracking sentiment towards news entities from Arabic news on social media. Future Generation Computer Systems. Elsevier B.V. https://doi.org/10.1016/j.future.2021.01.015

Aria, M, Cuccurullo, C. (2017) Bibliometrix: An R-tool for comprehensive science mapping analysis. Elsevier https://doi.org/10.1016/j.joi.2017.08.007

Azzouzi, M.E, Coatrieux, G, Bellafqira, R, Delamarre, D, Riou, C, Oubenali, N, Cabon, S, Cuggia, M, Bouzillé, G (2024), Automatic de-identification of French electronic health records: a cost-effective approach exploiting distant supervision and deep learning models. BMC Medical Informatics and Decision Making. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185310221&doi=10.1186

Bach,N.X., Thuy,N,T,T., Chien,D,B., Duy,T,K., Hien,T,M., Phuong,T,M. (2019). Reference Extraction from Vietnamese Legal Documents. The Tenth International Symposium on Information and Communication Technology. ACM. https://doi.org/10.1145/3368926.3369731

Bajaj, G, Kursuncu, U, Gaur, M, Lokala, U, Hyder, A, Parthasarathy, S, Sheth, A. (2022). Knowledge-Driven Drug-Use NamedEntity Recognition with Distant Supervision. Studies in Health Technology and Informatics. IOS Press BV. https://doi.org/10.3233/shti220048

Barachi, M.E., Mathew, S.S., Alkhatib, M. (2022). Combining Named Entity Recognition and Emotion Analysis of Tweets for Early Warning of Violent Actions. 2022 7th International Conference on Smart and Sustainable Technologies, SpliTech 2022. https://ieeexplore.ieee.org/document/9854231

Begen, P.N and Vidisaova, L. (2022). Development of an algorithm for fixing the citizens' assessments of digital transformation processes based on text analysis. ACM International Conference Proceeding Series. https://10.1145/3560107.3560203

Bouabdallaoui, I, Guerouate, F, Bouhaddour, S, Saadi, C, Sbihi, M. (2022). Named Entity Recognition applied on Moroccan tourism corpus. Procedia Computer Science. Elsevier B.V. https://doi.org/10.1016/j.procs.2021.12.256

Cardenas, P., Obara, B., Theodoropoulos, G., Kureshi, I. (2019). Defining an Alert Mechanism for Detecting Likely Threats to National Security. Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019. https://ieeexplore.ieee.org/document/8622569

Cenikj, G,Vitanova, B. Eftimov, T. (2021). Skills Named-Entity Recognition for Creating a Skill Inventory of Today's Workplace. Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021. https://ieeexplore.ieee.org/document/9671435

Chen, J, Huang, Y, Yang, F, Li, C. (2020). A novel named entity recognition approach of judicial case texts based on BiLSTM-CRF. 12th International Conference on Advanced Computational Intelligence, ICACI 2020. IEEE. https://ieeexplore.ieee.org/document/9177731

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. (2019). NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1, pp. 4171-4186. https://doi.org/10.48550/arXiv.1810.04805

Dogra, V, Singh, A, Verma, S, Alharbi, A, Alosaimi, W. (2021). Event study: Advanced machine learning and statistical technique for analyzing sustainability in banking stocks. MDPI. https://doi.org/10.3390/math9243319

Eva, M.B, Weber, N. (2022). Councils in Action: Automating the Curation of Municipal Governance Data for Research. Proceedings of the Association for Information Science and Technology. John Wiley and Sons Inc. https://doi.org/10.1002/pra2.601

Gangadharan, V, Gupta, D. (2020). Recognizing Named Entities in Agriculture Documents using LDA based Topic Modelling Techniques. Procedia Computer Science. Elsevier B.V. https://doi.org/10.1016/j.procs.2020.04.143

Garat, D, and Wonsever, D. (2022). Automatic Curation of Court Documents: Anonymizing Personal Data. Information (Switzerland). MDPI. https://doi.org/10.3390/info13010027

Gasmi, H., Laval, J., Bouras, A. (2019). Information extraction of cybersecurity concepts: An LSTM approach. Applied Sciences (Switzerland) MDPI, 9 (19), art. no. 3945. https://doi.org/10.3390/app9193945

Grishman, R and Sundheim, B. (1995). Design of the MUC-6 Evaluation. In Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995. https://doi.org/10.3115/1072399.1072401

Han, X, Wang, J (2019). Earthquake Information Extraction and Comparison from Different Sources Based on Web Text. Geo-Information. MDPI. https://doi.org/10.3390/ijgi8060252

He, S, Yang, H, Zheng, X, Wang, B, Zhou, Y, Xiong, Y, Zeng, D. (2019). Massive meme identification and popularity analysis in geopolitics. 2019 IEEE International Conference on Intelligence and Security Informatics, ISI 2019. IEEE. https://doi.org/10.1109/ISI.2019.8823294

Ivanin, V, Artemova, E, Batura, T, Ivanov, V, Sarkisyan, V, Tutubalina, E, Smurov, I. (2021). RuREBus: A Case Study of Joint Nameds Entity Recognition and Relation Extraction from E-Government Domain. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.48550/arXiv.2010.15939

Kulkarni, P., Deshmukh, V., & Rane, K. (2023). A framework for providing structured invoice document using optimized Bert enabled deep convolutional neural network classifier. 2023 Proceedings of the 7th International Conference on I-SMAC. IEEE. https://doi.org/10.1109/I-SMAC58438.2023.10290498

Lane, H., Nelson, C., & Sorgente, T. (2018). Named Entity Recognition with Python. O'Reilly Media, Inc. https://www.oreilly.com/library/view/natural-language-processing/9781787285101/ ch03s02.html

Li, J., Sun, A., Han, J., Li, C. (2022). A Survey on Deep Learning for Named Entity Recognition. IEEE Transactions on Knowledge and Data Engineering, 34 (1), pp. 50-70. https://doi.org/10.1109/TKDE.2020.2981314

Liu, Q, Wang, D, Zhou, M, Li, P, Qi, B, Wang, B. (2018). Chinese Governmental Named Entity Recognition. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Springer Verlag. http://dx.doi.org/10.1007/978-3-030-03520-4_2

Lutfi, S, Yasin, R, Barachi, M.E, Oroumchian, F, Imene, A, Samuel Mathew, S. (2021). Temporal behavioral analysis of extremists on social media: A machine learning based approach. 2021 6th International Conference on Smart and Sustainable Technologies, SpliTech 2021. IEEE. https://doi.org/10.23919/SpliTech52315.2021.9566446

Mansouri, A, Affendey, L.S, Mamat, A. (2008), Named Entity Recognition Approaches. Journal of Computer Science, vol. 8, no.2, pp. 339–344, 2008. https://www.researchgate.net/publication/ 238607553_Named_Entity_Recognition_Approaches

Martinez-Seis, B., Pichardo-Lagunas, L., Koff, H., Equihua, H., Perez-Maqueo, O., Hernandez-Huerta, A. (2022). Unified, Labeled, and Semi-Structured Database of Pre-Processed Mexican Laws. Data, MDPI. https://doi.org/10.3390/data7070091

Maurel, D, Morale, E, Thouvenin, N, Ringot, P, Turri, A. (2019). Istex: A database of twenty million scientific papers with a mining tool which uses named entities. Information (Switzerland). MDPI. https://doi.org/10.3390/info10050178

Nemes, L., Kiss, A. (2021). Information extraction and named entity recognition supported social media sentiment analysis during the COVID-19 pandemic. Applied Sciences (Switzerland), 11 (22), art. no. 11017. https://doi.org/10.3390/app112211017

Niu, H, Zeng, Z. (2018), A New Efficiency Approach for Chinese Litigants Extraction. Procedia Computer Science. Elsevier B.V. https://doi.org/10.1016/j.procs.2018.03.049

Park, J. S, Kim, G.W, Lee, D.H. (2020). Sensitive Data Identification in Structured Data through GenNER Model based on Text Generation and NER. ACM International Conference Proceeding Series. ACM. https://doi.org/10.1145/3398329.3398335

Pena, P, Aznar, R, Montanes, R, Del Hoyo, R. (2018). Open Data for Public Administration: Exploitation and semantic organization of institutional web content. Procesamiento del Lenguaje Natural. Sociedad Espanola para el Procesamiento del Lenguaje Natural. https://rua.ua.es/dspace/bitstream/10045/81360/1/PLN_61_21.pdf

Pimpisal, Simud, A, Sanglerdsinlapachai, T, Surasvadi,N, Plangprasopchok,N, Anon. (2021). Named Entity Recognition of Thai Documents using CRF with a Simple Data Masking Technique. 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing, iSAI-NLP 2021. https://doi.org/10.1109/iSAI-NLP54397.2021.9678156

PRISMA 2020. PRISMA Statement: Checklist. http://www.prisma-statement.org/PRISMAStatement/Checklist

Pugliese, D.P.L, Guerriero, F, Macrina, G, Messina, E. (2021). A Natural Language Processing Tool to Support the Electronic Invoicing Process in Italy. Proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021. https://doi.org/10.1109/IDAACS53288.2021.9660987

Ramdhani, T.W, and Budi, I, Purwandari, B. (2021). Optical Character Recognition Engines Performance Comparison in Information Extraction. International Journal of Advanced Computer Science and Applications. https://dx.doi.org/10.14569/IJACSA.2021.0120814

Sandescu, C, Dinisor, A, Vladescu, C.V, Grigorescu, O, Corlatescu, D, Dascalu, M, Rughinis, R. (2022). Extracting Exploits And Attack Vectors From Cybersecurity News Using NLP. UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science. Politechnica University of Bucharest. https://www.scientificbulletin.upb.ro/rev_docs_arhiva/ full90f_939135.pdf

Sensuse, D.I, Putro, P.A.W, Rachmawati, R., Sunindyo, W.D. (2022) Initial Cybersecurity Framework in the New Capital City of Indonesia: Factors, Objectives, and Technology. Information 2022, 13, 580. https://doi.org/10.3390/info13120580

Shen, Z, Spruit, M. (2021). Automatic extraction of adverse drug reactions from summary of product characteristics. Applied Sciences (Switzerland). MDPI. https://doi.org/10.3390/app11062663

Silvestri, S, Gargiulo, F, Ciampi, M. (2021). Iterative Annotation of Biomedical NER Corpora with Deep Neural Networks and Knowledge Bases. Applied Sciences (Switzerland). MDPI. https://doi.org/10.3390/app12125775

Street,M, Mestric,I,I., Ndoni,A., Lenk,P., Teufert,J., Figueiredo,N. (2022). Data Driven Decision Support during COVID. International Conference on Military Communications and Information Systems. Elsevier. https://doi.org/10.1016/j.procs.2022.09.013

Suat-Rojas, N, Gutierrez-Osorio, C, Pedraza, C. (2022). Extraction and Analysis of Social Networks Data to Detect Traffic Accidents. Information (Switzerland). MDPI. https://doi.org/10.3390/info13010026

Sufi, F.K. (2022). Identifying the drivers of negative news with sentiment, entity, and regression analysis. (2022) International Journal of Information Management Data Insights, 2 (1), art. no. 100074. https://doi.org/10.1016/j.jjimei.2022.100074

Sutphin, C, Lee, K, Yepes, A.J, Uzuner, Ö, McInnes, B.T. (2022). Adverse drug event detection using reason assignments in FDA drug labels. Journal of Biomedical Informatics. Academic Press Inc. https://doi.org/10.1016/j.jbi.2020.103552

Tan, F, Yang, S, Wu, X, Xu, J. (2020). Exploring the relation between biomedical entities and government funding. CEUR Workshop Proceedings. CEUR-WS. https://ceur-ws.org/Vol-2658/paper6.pdf

Toledo,C.V, Dijk,F.V., Spruit.M. (2020). Dutch Named Entity Recognition and De-identification Methods for the Human Resource Domain. International Journal on Natural Language Computing (IJNLC). https://doi.org/10.48550/arXiv.2106.02287

Wang, L, Li, S, Yan, Q, Zhou, G. (2018). Domain-specific named entity recognition with document-level optimization. ACM Transactions on Asian and Low-Resource Language Information Processing. ACM. https://doi.org/10.1145/3213544

Wang, Q, You, H. (2022). A Study on BNM-cBLSTM for Financial Sentiment Analysis in European Bond Markets Based on mpBC-ELMo. 2022 International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI). https://doi.org/10.1109/ICDACAI57211.2022.00032

Wang, Y, Wang, Z, Li, H, Huang W. (2023). Named Entity Recognition in Threat Intelligence Domain Based on Deep Learning. Journal of Northeastern University. https://doi.org/10.12068/j.issn.1005-3026.2023.01.005

White, N.M. (2022). The Hmong Medical Corpus: a biomedical corpus for a minority language. Language Resources and Evaluation. Springer Science and Business Media B.V. https://link.springer.com/content/pdf/10.1007/s10579-022-09596-2.pdf

Wicaksono, A.T, Mariyah, S. (2019). Social Network Analysis of Health Development in Indonesia. ICICOS 2019 - 3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings. IEEE. https://doi.org/10.1109/ICICoS48119.2019.8982482

Xiong, Z. and Kong, D. and Xia, Z. and Xue, Y. and Song, Z. and Wang, P. (2021). Chinese Government Official Document Named Entity Recognition Based on Albert. 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2021. https://doi.org/10.1109/ICCCBDA51879.2021.9442540

Yaozu, Y, Jiangen, Z. (2020). Constructing government procurement knowledge graph based on crawler data. Journal of Physics: Conference Series. IOP Publishing Ltd. https://iopscience.iop.org/article/10.1088/1742-6596/1693/1/012032/pdf

Zacharis, A., Gavrila, R., Patsakis, C., & Ikonomou, D. (2023). AI-assisted cyber security exercise content generation: Modeling a cyber conflict. International Conference on Cyber Conflict (CYCON), 2023-May, 217–238. https://doi.org/10.23919/CyCon58705.2023.10181930

Zhang, T, Liu, M, Ma, C, Tu, Z, Wang, Z. (2021). A Text Mining based Method for Policy Recommendation. Proceedings - 2021 IEEE International Conference on Services Computing, SCC 2021. IEEE. https://doi.org/10.1109/SCC53864.2021.00036