International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I7P142

Volume 72 Issue 7, 381-387, July 2024

Original Article

Utilizing Named Entity Recognition for Web-Based Resume Scoring

Ivic Jan A. Biol[1], Cristopher C. Abalorio[2]

[1,2] College of Computing and Information Sciences, Caraga State University, Butuan City, Philippines.

[2]Corresponding Author : ccabalorio@carsu.edu.ph

Received: 06 March 2024 Revised: 20 May 2024 Accepted: 02 July 2024 Published: 26 July 2024

Abstract - *The overwhelming number of job applicants received by companies has made resume evaluation a time-consuming task for recruiters. Online recruitment platforms have emerged as a solution for automating the matching of job openings with suitable resumes. This study analyzes the use of Named Entity Recognition (NER) to automate the evaluation of resumes in the hiring process. NER was employed as a resume scorer to extract relevant skills, education, and work experience from resumes. By identifying named entities, such as programming languages, education institutions, and job titles, the system efficiently assessed candidate qualifications and matched them to job requirements. This study utilized a dataset of 1,014 annotated resumes, and the RoBERTa NER model was fine-tuned using spacy transformers. In addition, the NER model for job descriptions was trained using a dataset of 200 job descriptions. The results demonstrated improvements in model performance over training epochs, with increased precision, recall, and F1 scores. This study highlights the potential of web-based resume scorers in automating resume evaluations and suggests directions for future research in this area*.

Keywords - *Natural Language Processing, Named Entity Recognition, Resume scorer, RoBERTa.*

1. Introduction

In the hiring process, the initial responsibility of a recruiter is to evaluate the resumes of job applicants. When a company has a job opening, it often receives an overwhelming number of emails from interested candidates on a daily basis. For recruiters, sifting through this large volume of applicants to identify potential candidates for the job can be a time-consuming and demanding task [1][2]. The increasing data pertaining to online recruitment has made job-resume matching increasingly important in the hiring process. Candidates' rankings are determined by the information provided in their resumes, including their education, work history, and skills. This means that their qualifications, past jobs, and abilities play a crucial role in determining their positions. In today's recruitment procedure, most organizations depend on internet-based platforms and services to receive resumes from job applicants. These online platforms and websites dedicated to resume submissions and searches are widely used globally [3][4]. They provide a convenient and centralized system for candidates to submit their profiles, making it easier for recruiters to access and review the applications [5][6][7]. However, existing systems often struggle with effectively and accurately parsing and ranking resumes due to the unstructured and varied nature of resume formats. This research identifies a significant gap in the current methodologies: the lack of a robust, automated mechanism to extract and evaluate key candidate information consistently and accurately. While some platforms employ basic keyword matching, this approach is insufficient for a nuanced understanding of a candidate's qualifications and suitability for a role. In this study, the researchers used Named Entity Recognition (NER). NER is employed as a resume scorer to automatically extract relevant skills, education, and work experience from the resumes. By identifying named entities, such as programming languages, education institutions, and job titles, the system can more efficiently assess the candidate's qualifications and match them to the job requirements. NER streamlines the evaluation process by automating the extraction of critical information, enabling recruiters to make informed decisions and saving time in reviewing resumes.

2. Literature Review

In this section, a literature review is presented related to the development of the Resume Scorer system utilizing Named Entity Recognition. This involves the thorough investigation of research within the field, focusing on the current state-of-the-art practices, which will serve as a foundation for this project.

2.1. Natural Language Processing

Natural Language Processing (NLP) is a vital field within computer science that extensively employs machine learning and computational linguistics [8][9]. Its primary objective is to facilitate easy and efficient interaction between humans and computers. In NLP, machines learn the syntax and semantics of human language, process it, and provide output to users [10]. NLP is a versatile tool; therefore it is not uncommon to find it being used in companies and organizations [11][12]. Taking this into consideration, a scoring system was developed to boost the efficiency of the selection process of employment using natural language processing. NLP has many branches, such as Text Classification [13][14], Sentiment Analysis [15][16], and Named Entity Recognition. Over the last few years, the employment of NLP methods has grown significantly thanks to the progress in deep learning systems like Transformers, which have completely transformed how languages are understood or created. It is these very models, for instance, BERT (Bi-directional Encoder Representations from Transformers) as well as GPT (Generative Pre-trained Transformer), that have introduced new standards in areas such as language modeling, question answering, and dialogue generation, thereby taking it further. The mingling between NLP and other domains, such as computer vision and data analytics, still fosters creativity in the industry by offering advanced ways of solving intricate challenges. This means that there will always be new developments within NLP to help improve various activities like human-to-computer communication processes, decision-making, and finding information, among others.

2.2. Named Entity Recognition

Data is often quite confusing and unstructured in our current era, making NLP a crucial part of data processing. Recognizing and labelling named entities from text is known as Name Entity Recognition (NER); it is one such process of NLP [17]. NER trivializes information retrieval as a model can be trained [18][19][20] for obtaining detailed information from textual data [21]. A web-based resume scorer was developed using NER due to its efficiency in information retrieval.

The utilization of NER in various fields has been highly successful. For instance, in the health sector, NER is implemented to isolate and classify user data contained in clinic records, hence enhancing effective patient treatment and research [22][23]. Equally, for instance, in finance [24], NER plays a critical role in the identification as well as tracing of entities, including firms, stock prices, and financial indicators manifested in news items or reports, hence making it possible to make more enlightened decisions. NER's flexibility extends to how it can be integrated with other NLP tasks. For example, when NER is added to sentiment analysis [25][26], it can help us know how people feel about particular entities like products on sale or famous persons through social media messages, among other things. One last example would be question-answering systems; they need accurate identification of entities so that they give correct answers [27][28]. No wonder NER is often referred to as a game changer in natural language processing.

3. Materials and Methods

The following subsections present an overview of the essential elements necessary for this study utilizing a variant of the BERT model – RoBERTa for analyzing data. These elements encompass the data, research design, techniques for collecting and pre-processing data, machine learning models, metrics for evaluation, and the simulation results for the web application employed for model testing. Collectively, these components establish an inclusive framework in order to conduct meticulous and organized research, leveraging models to derive understandings from data and assess the performance of the model. Figure 1 illustrates the implementation and its research framework, offering a clear depiction of the study's adopted methodology.

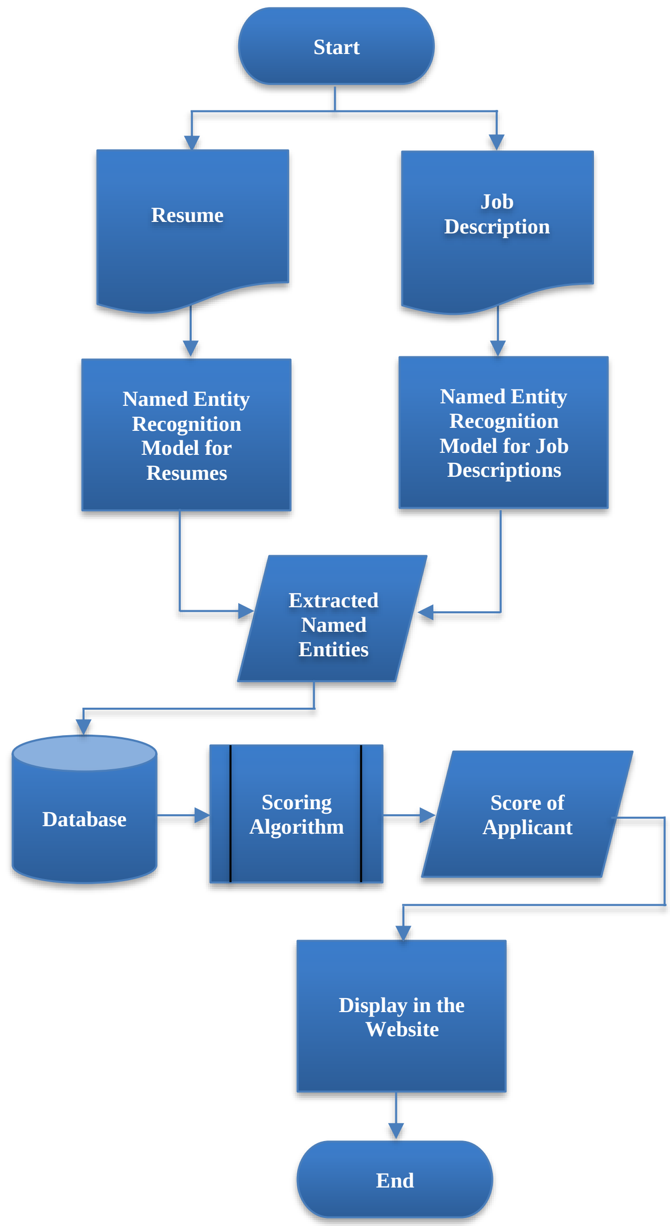

Fig. 1 Framework of NER Web-Based resume scoring

The elements of the study's framework include:

3.1. NER Model for Resumes

The dataset used for the model was created and annotated by Roman Shilpakar, and it contained 1,014 resumes obtained from various sources on the internet. The categories in the dataset were as follows: NAME, EMAIL ADDRESS, LOCATION, COLLEGE NAME, DEGREE, YEAR OF GRADUATION, UNIVERSITY, SKILLS, WORKED AS, COMPANIES WORKED AT, YEARS OF EXPERIENCE, DESIGNATION, LANGUAGE, CERTIFICATION, AWARDS, LINKEDIN LINK, CONTACT, and Unlabelled.

Shilpakar then used the dataset to finetune the RoBERTa NER model through spacy transformers, a Python library for natural language processing tasks.

3.2. NER Model for Job Descriptions

The model was trained using a dataset containing 200 job descriptions obtained from the internet. It was created and annotated by Roman Shilpakar with the following entity labels: CERTIFICATION, DEGREE, EXPERIENCE, JOBPOST, SKILLS. The dataset was then used to fine-tune the RoBERTa NER model.

3.3. Preprocessing

This study used preprocessing techniques to improve the quality of text input, including converting all text input into smaller cases to ensure uniformity in the treatment of words, splitting text into words or sub-words, getting rid of such punctuation marks that may interfere with named entities but are not needed themselves and finally eliminating commonly used words during processing.

3.4. RoBERTa

Following the creation and annotation of the datasets, the training process was facilitated using a configuration file, which specified the settings and parameters for the different components of the NER model. These components include:

- Paths – specifies the paths to essential files, including training data, development data, pre-trained word vectors, and initialization token-to-vector mappings.
- System – defines system-level settings, including the GPU allocator and the random seed for reproducibility.
- NLP – includes NLP-specific settings, such as the language, the pipeline components (Transformer and NER), and batch size.
- Components – defines the components used in the pipeline and their corresponding configurations, including the NER component and the transformer component.
- Corpora – specifies the corpora used for training and evaluation, including details such as paths to the data and additional settings.
- Training – encompasses the training settings, including gradient accumulation, corpora for training and evaluation, random seed, optimizer configurations, and scoring weights for evaluation.

The configuration file was utilized in the training process. It involved multiple iterations with a batch size of 128 and a test size of 0.3. The optimizer used for training was Adam, with specific settings such as beta values, L2 weight decay, gradient clipping, and a learning rate schedule. An early stopping was employed to monitor the model's performance on the training dataset. The evaluation frequency was set to every 200 iterations, and the scoring weights for evaluation metrics were configured accordingly.

3.5. Evaluation Method

The following metrics were used to evaluate the performance of the NER model:

Entity Precision – evaluates the percentage of accurately predicted entities relative to all entities predicted by the model.

$$Entity\ Precision = \frac{True\ Positive}{True\ Positive + False\ Postive} \quad (1)$$

Entity Recall – evaluates the ratio of correctly predicted entities among all actual entities present in the dataset.

$$Entity\ Recall = \frac{True\ Positive}{True\ Positive + False\ Postive} \quad (2)$$

Entity F1 Score – the balanced metric of a model's performance, calculated as the harmonic mean of precision and recall, merging both aspects into a singular value.

$$Entity\ F1 = 2 * \frac{Entity\ Precision * Entity\ Recall}{Entity\ Precisio + Entity\ Recall} \quad (3)$$

Loss NER – also known as cross-entropy loss, measures the discrepancy between the predicted named entity labels and the true labels.

$$Cross\ Entropy\ Loss = -\sum_{\substack{i \\ y_i \neq 0}} Log(\bar{y}_i) \quad (4)$$

Where y_i represents the true label, and $\bar{y}$ is the predicted label.

3.6. Scoring Process

The researchers conducted an interview with a representative from the Human Resources department of ACLC College of Butuan to determine which sections of a resume they give more importance to when selecting a candidate for employment. As a result of the interview, the researchers found out that the HR department places the most importance on the competency of the applicant, secondly, the previous work job, and lastly, the character reference. The researchers founded a system that rates and balances factors following the under-mentioned guidelines:

- Competence (0.65) – covers the education level of an applicant and their abilities.
- Working experience (0.20) – prior work positions at the given place.
- Referee opinions (0.15) – consultant reports provided by potential workers that testify their talents and morality.

This is a rating system where all these factors are converted into numbers.

4. Results and Discussion

4.1. Results

The following section describes the results from training and testing the NER model results and discussion. Table 1 presents the results from training the Resume NER Model. Initially, during the first epoch, the loss value is relatively high (1411.23), indicating poor performance. The F1 score is also quite low (0.08), suggesting low accuracy in identifying entities. The precision (0.04) and recall (0.30) are also suboptimal. As training progresses, we see improvements in the model's performance. The loss value decreases significantly, and the F1 score, precision, and recall increase gradually. The model continued to improve until the 20th epoch, where the scores remained consistent until the 44th epoch, where it had the final F1 score of 85.42, a precision of 84.25, a recall score of 86.63, and a loss value of 1997.00.

Table 2 presents the training results of the Job Description NER Model. At the beginning (epoch 0), the loss value is relatively high (834.01), indicating poor performance. The F1 score, precision, and recall are all 0, suggesting that the model is not identifying any entities correctly. As training progresses, we see improvements in the model's performance. The loss value decreases, and the F1 score, precision, and recall values increase. From the 19th epoch onward, the model's performance shows some fluctuations, but the F1 score, precision, and recall values generally increase gradually. Towards the end of training (at epoch 50), the model achieved an F1 score of 57.75, a precision score of 56.72 and a recall of 58.81 values, which also showed some improvement.

Figure 2 illustrates a visual representation of the results from training the Resume NER Model, while Figure 3 illustrates the Job Description NER Model using a line graph. At approximately 50,000 at epoch 0, the blue line indicates a loss metric that plummets steeply to below 10,000 within epoch 20. This rapid decrease indicates effective initial learning and parameter adjustment.

In addition, the loss declines further, although not as much as it did before at about epoch 20, suggesting convergence and stabilization. Table 3 displays the text from a resume and job description as well as the named entities extracted from them. From the left is the text obtained from a resume, and next to it are the extracted named entities, which are grouped into 7 different categories. Further to the right, the text from a job description, as well as the extracted entities, can be seen. These entities are grouped into 4 categories.

4.2. Discussions

The researchers have developed a web-based resume-scoring system that utilizes Named Entity Recognition (NER) and the RoBERTa language model. This system can evaluate resumes by identifying entities such as awards, certifications, college names, work experience, skills, and more.

Two models were used for this purpose: the resume NER model and the Job Description NER model. During the training process, both models initially showed low performance, with an F1 score, precision, and recall of 0 in the first epoch. However, as the training progressed, both models gradually improved their performance, resulting in high F1 scores, precision, and recall values, as seen in Tables 1 and 2. Both models show a sharp reduction in loss function, which is a good sign of learning. Poor initial performance on the metrics of precision, recall and F1-score implies that there could have been training difficulties experienced at early stages. The improvement over epochs reveals the model’s capacity to learn from data and improve its entity recognition abilities.

Figures 2 and 3 visually show remarkable improvements for both models that vividly showcase their improved performance. Also, Table 3 gives an account of what has been obtained from resume content as well as job description content and presents named entities identified in 7 groups for the resume and 4 groups for the job description. These two examples demonstrate not only how effective these techniques are but give a more detailed analysis of this fact.

Table 1. Training results of the resume NER model

Epoch	LOSS NER	ENTS F1	ENTS P	ENTS R
0	1411.23	0.08	0.04	0.30
4	48336.09	0.27	0.82	0.16
8	23833.26	74.34	74.23	74.45
12	9191.00	83.48	81.75	85.29
16	6350.63	85.15	84.06	86.28
20	5192.89	85.51	84.76	86.28
24	4481.52	85.07	83.63	86.56
28	3427.31	84.73	83.54	85.95
32	2430.41	85.50	83.96	87.11
36	2733.66	86.14	85.46	86.83
40	1979.00	85.77	84.94	86.62
44	1997.08	85.42	84.25	86.63

Table 2. Training results of the job description NER model

Epoch	LOSS NER	ENTS F1	ENTS P	ENTS R
0	834.01	0.00	0.00	0.00
3	49393.11	23.78	46.08	16.03
7	13718.34	42.14	58.74	32.85
11	6560.12	54.42	53.57	55.29
15	3427.80	54.43	54.69	54.17
19	2243.77	58.30	58.92	57.69
23	2100.16	50.80	41.22	66.19
26	1559.77	57.52	58.67	56.41
30	955.34	55.88	58.26	53.69
34	954.91	57.60	57.84	57.37
38	832.25	57.23	53.48	61.54
42	852.09	57.56	54.96	60.42
46	899.00	55.38	51.59	59.78
50	888.25	57.75	56.72	58.81

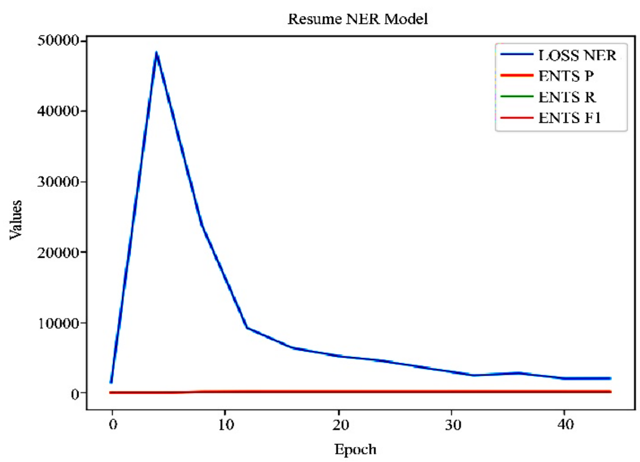

Fig. 2 Line graphs of the resume NER model

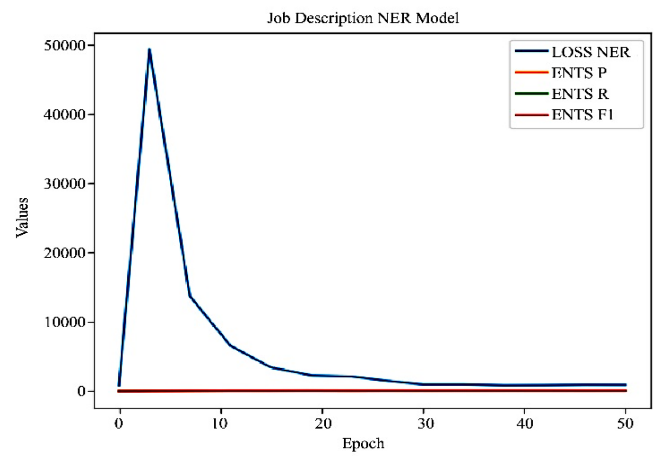

Fig. 3 Line graphs of the job description NER model

Table 3. Samples of Text and Extracted Entities from Resume and Job Description

Resume Text	Resume Entities	Job Description Entities
Contact www.linkedin.com/in/shreya- bhandari-a0094816b (LinkedIn) shreyaabh.pythonanywhere.com/ (Portfolio) Top Skills Writing Python (Programming Language) Django Languages Nepali (Native or Bilingual) English Hindi Shreya Bhandari Python Developer|AI/ML enthusiast| Graphic Designer Kathmandu Experience Danson Solutions Business Analyst April 2022 - Present (4 months) Girls in Tech - Nepal 7 months Event Team Ambassador April 2022 - Present (4 months) Kathmandu, Bāgmatī, Nepal Support Squad Ambassador January 2022 - April 2022 (4 months) Mystery Gift Nepal Business Owner February 2022 - Present (6 months) Kathmandu, Bāgmatī, Nepal Danson Training Project Lead June 2021 - Present (1 year 2 months) United States Finnove Technologies 6 months Associate Software Engineer November 2021 - November 2021 (1 month) Kathmandu, Bāgmatī, Nepal Python Developer June 2021 - November 2021 (6 months) Kathmandu, Bāgmatī, Nepal Danson Solutions Data Science Intern Page 1 of 2 August 2021 - October 2021 (3 months) United States Chimp Vine Graphic Designer March 2021 - July 2021 (5 months) United States Vertex Report Content Writer June 2020 - June 2021 (1 year 1 month) Nepal Vedic Himalayan Adve Content Writer March 2017 - December 2019 (2 years 10 months) Kantipath, Kathmandu Education Nepal College of Information Technology Bachelor of Engineering - BE, Software engineering Kathmandu Model College High School , Science (Physical Group) Meridian International School SLC Page 2 of 2	'COMPANIES WORKED AT': ['Finnove Technologies 6 months'], 'LANGUAGE': ['Nepali', 'English', 'Hindi'], 'LINKEDIN LINK': ['www.linkedin.com/in/shreya-', 'bhandari', '-', 'a0094816b'], 'NAME': ['Shreya Bhandari Python Developer|AI/'], 'SKILLS': ['Python', 'Django'], 'WORKED AS': ['Associate Software Engineer'], 'YEARS OF EXPERIENCE': ['4 months', '4 months', '6 months', '1 year 2 months', '1 month', '6 months', '3 months', '5 months', '1 year 1 month', '2 years 10 months']	'DEGREE': ['Diploma in Computer Science (College or University)'], 'EXPERIENCE': ['1 years'], 'JOBPOST': ['Python Developer'], 'SKILLS': ['Django', 'Python', 'Bootstrap', 'HTML', 'Jquery', 'CSS', 'Ajax', 'Javascript', 'Bootstrap', 'GIT', 'lab', 'Docker', 'Kubernetes', 'OpenShift', 'Vue.js', 'Angular', 'React', 'HTML5', 'CSS3', 'GIT', 'Webpack', 'NPM']

5. Conclusion

Researchers have successfully developed web-based systems to address the inconvenience and time-consuming nature of evaluating job applicants' resumes. The Resume NER Model achieved a final F1 score of 85.42, indicating a good balance between precision (84.25) and recall (86.63). This demonstrates the model's proficiency in accurately identifying entities, such as awards, certifications, college names, work experience, and skills within resumes. On the other hand, the Job Description NER Model obtained a final F1 score of 57.75, a precision score of 56.72 and a recall score of 58.81. Although this model exhibited a relatively lower performance than the Resume NER Model, it still demonstrated the ability to identify entities within job descriptions, with potential for improvement. In conclusion, the Web-Based Resume Scorer shows promise in automating the resume evaluation process by effectively extracting crucial information. The results highlight the potential for further enhancements and refinements to improve the precision, recall, and overall performance in assessing resumes and matching them with job descriptions.

References

[1] Sujit Amin et al., "Web Application for Screening Resume," *2019 International Conference on Nascent Technologies in Engineering*, Navi Mumbai, India, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Matheus Werner, and Eduardo Laber, "Extracting Section Structure from Resumes in Brazilian Portuguese," *Expert Systems with Applications*, vol. 242, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ashif Mohamed et al., "Smart Talents Recruiter - Resume Ranking and Recommendation System," *2018 IEEE International Conference on Information and Automation for Sustainability*, Colombo, Sri Lanka, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Agnieszka Wosiak, "Automated Extraction of Information from Polish Resume Documents in the IT Recruitment Process," *Procedia Computer Science*, vol. 192, pp. 2432-2439, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Shuqing Bian et al., "Learning to Match Jobs with Resumes from Sparse Interaction Data Using Multi-View Co-Teaching Network," *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 65-74, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Girish K. Palshikar et al., "RINX: A System for Information and Knowledge Extraction from Resumes," *Data & Knowledge Engineering*, vol. 147, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Shun Luo, and Juan Yu, "ESGNet: A Multimodal Network Model Incorporating Entity Semantic Graphs for Information Extraction from Chinese Resumes," *Information Processing & Management*, vol. 61, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Rushan Geng et al., "Planarized Sentence Representation for Nested Named Entity Recognition," *Information Processing & Management*, vol. 60, no. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Pu Li et al., "EPIC: An Epidemiological Investigation of COVID-19 Dataset for Chinese Named Entity Recognition," *Information Processing & Management*, vol. 61, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Aditya Jain, Gandhar Kulkarni, and Vraj Shah, "Natural Language Processing," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 1, pp. 161-167, 2018. [CrossRef] [Publisher Link]

[11] Alexander Smirnov et al., "Natural Language Processing Workflow for Customer Request Analysis in a Company," *IFAC-PapersOnLine*, vol. 54, no. 1, pp. 1206-1211, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ángeles Aldunate et al., "Understanding Customer Satisfaction via Deep Learning and Natural Language Processing," *Expert Systems with Applications*, vol. 209, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Rasmita Rautray, and Rakesh Chandra Balabantaray, "An Evolutionary Framework for Multi Document Summarization Using Cuckoo Search Approach: MDSCSA," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 134-144, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Saman Jamshidi et al., "Effective Text Classification Using BERT, MTM LSTM, and DT," *Data & Knowledge Engineering*, vol. 151, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Vicente A. Pitogo, and Christine Diane L. Ramos, "Social Media Enabled e-Participation: A Lexicon-Based Sentiment Analysis Using Unsupervised Machine Learning," *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, Athens Greece, pp. 518-528, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yanying Mao, Qun Liu, and Yu Zhang, "Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Basra Jehangir, Saravanan Radhakrishnan, and Rahul Agarwal, "A Survey on Named Entity Recognition-Datasets, Tools, and Methodologies," *Natural Language Processing Journal*, vol. 3, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Qiqi Chen et al., "CareerMiner: Automatic Extraction of Professional Network from Large Chinese Resume Data," *Franklin Open*, vol. 6, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yu-Chun Wang, Richard Tzong-Han Tsai, and Wen-Lian Hsu, "Web-Based Pattern Learning for Named Entity Translation in Korean-Chinese Cross-Language Information Retrieval," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3990-3995, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[20] Hongjin Kim, and Harksoo Kim, "Recursive Label Attention Network for Nested Named Entity Recognition," *Expert Systems with Applications*, vol. 249, no. B, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Desislava Petkova, and W. Bruce Croft, "Proximity-Based Document Representation for Named Entity Retrieval," *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, Lisbon Portugal, pp. 731-740, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[22] Diego Pinheiro da Silva et al., "Exploring Named Entity Recognition and Relation Extraction for Ontology and Medical Records Integration," *Informatics in Medicine Unlocked*, vol. 43, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Aleksandar Kaplar et al., "Evaluation of Clinical Named Entity Recognition Methods for Serbian Electronic Health Records," *International Journal of Medical Informatics*, vol. 164, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Han Zhang et al., "Chinese Named Entity Recognition Method for the Finance Domain Based on Enhanced Features and Pretrained

Language Models," *Information Sciences*, vol. 625, pp. 385-400, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Fahim K Sufi, "Identifying the Drivers of Negative News with Sentiment, Entity and Regression Analysis," *International Journal of Information Management Data Insights*, vol. 2, no. 1, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Manman Luo, and Xiangming Mu, "Entity Sentiment Analysis in the News: A Case Study Based on Negative Sentiment Smoothing Model (NSSM)," *International Journal of Information Management Data Insights*, vol. 2, no. 1, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Ze Hu, and Xiaoning Ma, "A Novel Neural Network Model Fusion Approach for Improving Medical Named Entity Recognition in Online Health Expert Question-Answering Services," *Expert Systems with Applications*, vol. 223, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Fang Wang et al., "Named Entity Disambiguation for Questions in Community Question Answering," *Knowledge-Based Systems*, vol. 126, pp. 68-77, 2017. [CrossRef] [Google Scholar] [Publisher Link]