e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

EVALUATION METHODOLOGIES AND PERFORMANCE METRICS IN SUPERVISED NER FOR JOURNAL ARTICLES: A CRITICAL ANALYSIS

Chakradhar Reddy Peddavenkatagari*1

*1Student, Networking and communications, SRMIST, India.

ABSTRACT

Named Entity Recognition (NER) holds paramount importance in the realm of natural language processing, especially within the context of journal articles. This paper conducts an exhaustive analysis of supervised learning approaches employed for NER in journal articles, highlighting the nuances of evaluation methodologies and the criticality of pivotal performance metrics.

The research scrutinizes evaluation metrics such as precision, recall, and the F1-score, elucidating their role in assessing model accuracy at both token and entity granularity levels. Additionally, the study delves into the deployment of cross-validation techniques, vital for bolstering the resilience and generalizability of NER models across heterogeneous datasets.

The paper accentuates the significance of juxtaposing results against baseline models to discern the efficacy and pinpoint areas ripe for enhancement within supervised learning paradigms. Error analysis is identified as an instrumental phase, enabling the detection of recurrent error patterns and guiding targeted model optimizations. Furthermore, the paper underscores the imperative of assessing model generalizability to novel data, illuminating the pragmatic viability of supervised learning techniques in real-world scenarios.

By providing a comprehensive overview of evaluation methodologies and performance considerations pertinent to NER in journal articles, this review aspires to arm researchers and professionals with invaluable insights, catalyzing advancements in both theoretical natural language processing research and its tangible applications.

Keywords- Deep Learning, Evaluation Metrics, Multi-task Learning, Named Entity Recognition, Natural Language Processing, Supervised Learning, Transfer Learning.

I. INTRODUCTION

Named Entity Recognition (NER) is a pivotal technique within the domain of Natural Language Processing (NLP) designed to identify and categorize entities in text. The primary objective of NER is to sift through unstructured textual data, enabling machines to comprehend and categorize pertinent entities. This capability finds applications across a plethora of domains including text mining, knowledge mapping, test generation, and knowledge graph creation, among others. Often referred to as entity extraction or entity fragmentation and recognition, NER is entrenched in various facets of Artificial Intelligence (AI) encompassing Machine Learning (ML), Deep Learning, and Neural Networks.

NER serves as an indispensable module in NLP frameworks, augmenting the functionality of tools like chatbots, sentiment analysis systems, and search engines. Its ubiquity extends across diverse sectors such as healthcare, finance, human resources (HR), customer support, higher education, and social analytics. By adeptly identifying, classifying, and extracting salient information from voluminous unstructured documents, NER obviates the necessity for laborious manual searches, thereby expediting the data extraction process. The potency of NER models in amplifying the capabilities of artificial intelligence is noteworthy.

These models bolster AI's proficiency in deciphering human language across applications like content analysis, translation, and text analytics. NER's underlying grammar leverages algorithms rooted in NLP and predictive modeling. These algorithms are meticulously trained on annotated datasets, where entities are pre-defined under specific categories such as people (PER), locations (LOC), organizations (ORG), expressions, percentages, financial values, and more. In essence, NER serves as a linchpin in facilitating effective data extraction and interpretation from textual data, thereby driving advancements in artificial intelligence and enhancing the analytical prowess of AI systems across various linguistic domains.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

II. LITERATURE SURVEY

The evolution and development of Named Entity Recognition (NER) have been extensively studied over the years, with notable contributions from researchers across various domains. Nadeau and Sekine's seminal study in 2007 laid foundational insights into NER systems, encompassing supervised, semi-supervised, and unsupervised methodologies. Their research delineated the prevailing characteristics of NER systems at that juncture and shed light on systems that continue to be operational today. Subsequent research by Sharnagat in 2014 delved deeper into the realm of NER, encapsulating advancements in supervised, semi-supervised, and unsupervised NER paradigms, along with pioneering forays into neural network-based NER systems.

A plethora of studies have since been undertaken, each focusing on specific names and linguistic domains. Notable mentions include biomedical NER by Leaman and Gonzalez (2008), Chinese medical NER by Lei et al. (2013), Arabic NER by Shaalan (2014) and Etaiwi et al. (2014), and NER tailored for Indian languages by Patil et al. (2016).

The predominant trajectory of contemporary NER research predominantly revolves around machine learning engineering models, encompassing supervised, semi-supervised, and unsupervised paradigms. Typically, these studies are inclined towards a singular linguistic domain or a specific sector. Interestingly, comprehensive studies elucidating modern neural network-based NER systems or comparative analyses across diverse languages and domains remain conspicuously sparse.In an endeavor to bridge this research gap, a comprehensive survey was undertaken, leveraging multiple search platforms including Google, Google Scholar, and Semantic Scholar.

The survey was meticulously designed to scrutinize articles encompassing various facets of NER, ranging from name recognition and neural architectures for NER to deep learning models tailored for NER. Each article was subjected to rigorous evaluation based on its relevance, novelty, and contribution to the NER discourse.

The evaluation process culminated in the review of 154 articles, out of which 83 were deemed pertinent for the research.

The selected articles provided invaluable insights into the contemporary landscape of NER, offering nuanced perspectives on neural architectures, feature engineering, and the best-performing models across diverse linguistic domains and application areas.

III. OBJECTIVES

Named Entity Recognition (NER) serves as a pivotal component in the realm of Natural Language Processing (NLP), facilitating the transition from unstructured textual data to structured information. The primary objective of NER is to sift through copious amounts of text, pinpoint specific segments, and categorize them based on predefined namespaces. This categorization process enables the conversion of unwieldy textual data into streamlined data structures, thereby enhancing the efficiency and efficacy of subsequent data-centric tasks such as data analysis, retrieval, and information extraction. By acting as a conduit between unstructured and structured data, NER empowers machines to traverse vast datasets, discern pertinent information, and present it in a categorized format.

This transformative capability of NER revolutionizes the data processing landscape, facilitating the seamless identification and extraction of salient information from voluminous data repositories.

In the context of NLP, NER plays a pivotal role in identifying various types of named entities such as personal names, organizational entities, locations, medical terms, numerical values, and financial metrics, among others. This granular categorization process categorizes the extracted data into predefined classes, thereby enabling a more nuanced understanding and interpretation of the underlying textual content. Understanding the intricacies of NER is paramount across a myriad of applications in NLP, as named entities often encapsulate the crux of the information embedded within textual data.

Whether it's deciphering medical records, parsing financial documents, or extracting actionable insights from academic literature, NER serves as an indispensable tool that augments the data processing capabilities, facilitates informed decision-making, and drives innovations across diverse domains.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

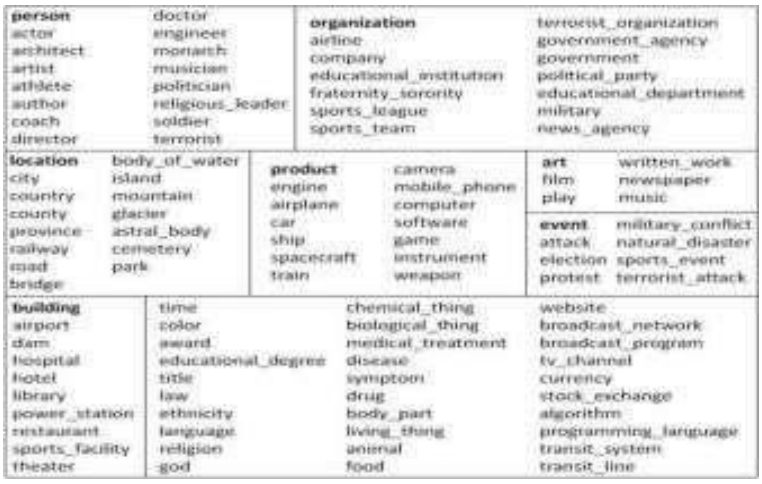
IV. METHODOLOGY

SUPERVISED METHOD

Supervised learning algorithms represent a category of machine learning techniques that operate by learning patterns and relationships from labeled training data. In the context of Named Entity Recognition (NER), supervised learning algorithms play a pivotal role in training models to accurately identify and classify named entities within textual data. Several sophisticated algorithms are employed in supervised learning for NER, each with its unique approach and strengths.

3.1 Hidden Markov Models

HMM is the first model developed by Bikel et al. It was applied to solve the NER problem. (1999) English version.Bikel introduced his IdentityFinder, a system for detecting NER.



Bikel's formulation of the Named Entity Recognition (NER) system problem is based on the premise that each word in a given context can be assigned only one name or label. This formulation is captured through the utilization of Hidden Markov Models (HMMs), which are generative probabilistic models adept at modeling sequences of observations. In Bikel's Identifier system, each word in the input sequence is assigned a label from a predefined set of classes or a "NOT-A-NAME" label if the word does not correspond to any desired class. The objective is to determine the most probable sequence of name classes (NC) given an observed sequence of words (W), which can be formulated as:

maxPr(NC|W)

Here,

Pr(NC|W)

Pr(NC|W) denotes the conditional probability of the name classes (NC) given the observed word sequence (W).

$$\Pr(\mathrm{NC}|\mathrm{W}) = \frac{\Pr(\mathrm{W}, \mathrm{NC})}{\Pr(\mathrm{W})}$$

Where:

Pr(W, NC) , Pr(W, NC) represents the joint probability of observing the word sequence (W) and the corresponding name classes (NC).

Pr(W), Pr(W) denotes the marginal probability of the observed word sequence (W).

HMMs, being generative models, aim to model the joint distribution Pr(W, NC) ,Pr(W, NC) and subsequently derive Pr(NC|W), Pr(NC|W) to identify the most likely sequence of name classes

given the observed word sequence. The state diagram for this HMM-based NER model delineates the transitions between different states corresponding to the observed words and the assigned name classes, facilitating the computation of the desired probabilities and the identification of the optimal sequence of name classes.In essence, Bikel's Identifier system leverages the power of HMMs to generate data (word sequence W and label NC) based on the distribution parameters, thereby enabling the extraction of named entities by determining the most probable sequence of name classes for a given sequence of words.

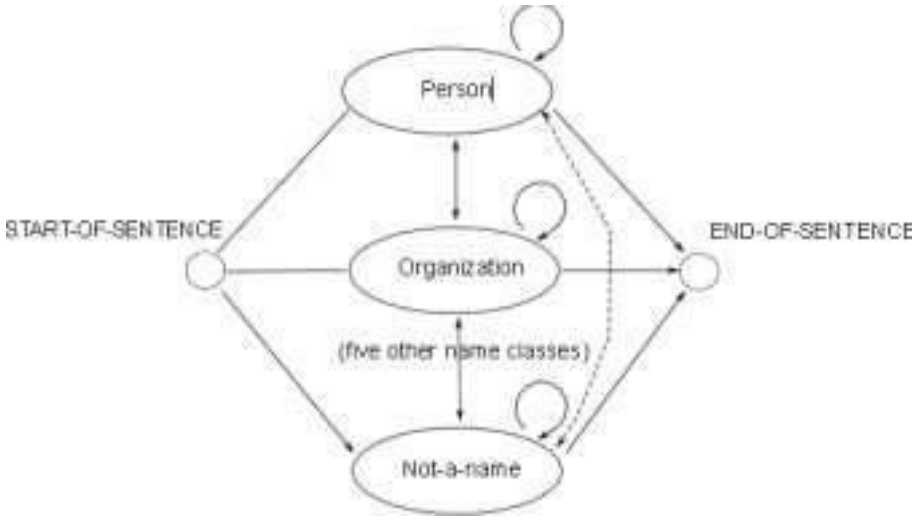
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Using the Viterbi algorithm Forney (1973), Pr(W, NC). -Maximize class allocation. Bikel modeled a generationon her three steps:

- Select name class nc depending on previous name class and word.
- Create first word in name class considering current name and previous name

3.2 Maximum Entropy Based Models

Maximum entropy models, unlike HMMs, are discriminative models. Using a set of features and training data, the model directly learns to weight discriminative features for classification. The goal of maximum entropy models is to maximize the entropy of the data to generalize as much of the training data as possible. In the ME model, each feature is associated with a parameter λi . Therefore, the conditional probability is obtained as follows:



$$P(f|h) = \frac{\prod_i \lambda_i^{g_i(h,f)}}{Z_\lambda(h)}$$

$$Z_\lambda(h) = \sum_f \prod_i \lambda_i^{g_i(h,f)}$$

The Maximum Entropy (MaxEnt) principle aims to maximize the entropy of the model while satisfying the constraints imposed by the training data. This ensures that the expected value of each feature in the MaxEnt model matches the empirical expectation of observed in the training corpus.

By doing so, the model captures the most unbiased, generalized representation of the training data, making it more robust and capable of generalizing well to unseen data.

The Viterbi algorithm is a dynamic programming algorithm used to find the most probable sequence of hidden states (or labels) given a sequence of observations (or features) in a Hidden Markov Model (HMM) or other probabilistic graphical models. In the context of Named Entity Recognition (NER), the Viterbi algorithm can be applied to find the most probable sequence of named entity labels for a given sequence of words, based on the probabilities computed by the MaxEnt model. Borthwick's MENE system (Borthwick, 1999) is an example of a comprehensive NER system that leverages multiple types of information and features to improve the accuracy of named entity recognition. The system utilizes various dictionaries, including a comprehensive dictionary and specialized single-word dictionaries (e.g., name, company name, company success), to enhance the recognition of named entities in different domains.The MENE system employs a wide range of features, such as binary properties, lexical properties, partial properties, external system output, compatibility, and problem-solving capabilities, to capture the diverse characteristics and contexts of named entities in text. By incorporating these features, the system aims to improve the performance and robustness of the MaxEnt model in identifying and classifying named entities accurately.

Curran's ME Tagger

Curran and Clark (2003) used the maximum entropy model for the naming problem. They use the softmax method to generate the probability P(y|x). Tagger uses a pattern of the form:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

P (y|x) =1 exp(Σλ f (x, y))

where y is the tag, x is the context and fi(x, y) is the feature with associated weight λi.

Hence the overall probability for the complete sequence of y1 • • • yn and words sequence w1 • •

wn is approximated as: n

P (y1 · · · yn|w1 · · ·wn) ≈ Pr(yi|xi) i=1

where xi is a context vector for each word wi. The tagger uses ray detection to find the most likely result for a sentence. Curran reported 84.89% accuracy for English test materials and 68.48% accuracy for German test materials for the CoNLL-2003 joint task.

3.3 SVM Based Models

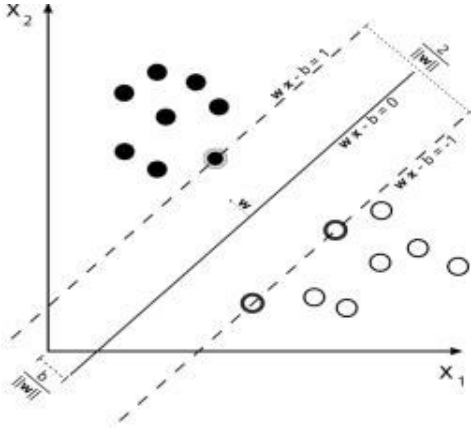
Support Vector Machine (SVM) is a powerful supervised learning algorithm introduced by Cortes and Vapnik in 1995. SVM is primarily designed for binary classification tasks, aiming to find the optimal hyperplane that separates the data points of different classes with the maximum margin. The points closest to the hyperplane on both sides are referred to as support vectors, which play a crucial role in defining the decision boundary and determining the classification of new, unseen data points.In the context of Named Entity Recognition (NER), SVM can be applied as a linear classifier to distinguish between named entities and non-named entities based on the features extracted from the text. The SVM model is characterized by two main parameters: the weight vector (W) perpendicular to the separating hyperplane and the bias term (b) that accounts for the deviation of the hyperplane from the origin.

The decision function of the SVM classifier can be defined as:

[f(x) = W cdot x + b]

If \(f(x) > 0), the sample (x) is classified as a positive sample (e.g., a named entity), otherwise it is classified as a negative sample (e.g., a non-named entity).In cases where the data points are not linearly separable, SVM can utilize a kernel trick to map the input features into a higher-dimensional space where the data becomes linearly separable, thereby accommodating some error in the distribution and preventing the classifier from overfitting the data.When dealing with multi-class classification problems, such as identifying named entities belonging to different categories (e.g., people, organizations, places), SVM can be extended using a one-vs-all (OvA) or one-vs-one (OvO) strategy. In the study by McNamee and Mayfield (2002), eight binary classifiers were trained to handle the multi-class NER task, where each classifier corresponds to one category and is responsible for distinguishing between that category and all other categories.

The feature set used in the SVM model typically includes a combination of character n-grams, word embeddings, and other linguistic features extracted from the text. In the mentioned study, a total of 258 characters and symbols and 1000 related words were used as features, resulting in a feature space size of 8806.For labeling the named entities, the CoNLL 2002 dataset was used, which consists of Spanish and Dutch language data. The reported F1 scores for the SVM-based NER system on the Spanish and Dutch datasets were 60.97 and 59.52, respectively, demonstrating the effectiveness of SVM in handling NER tasks across different languages and datasets.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

3.4 CRF Based Model

Conditional Random Fields (CRF) is a probabilistic graphical model that is widely used in pattern recognition and machine learning tasks, including Named Entity Recognition (NER). Introduced by Lafferty et al. in 2001, CRFs provide a framework for modeling the dependencies between observed and hidden variables in structured data, such as sequences or graphs.

In the context of NER, CRFs offer a principled approach to incorporate various features extracted from the text, such as word identities, part-of-speech tags, and contextual information, to predict the most likely sequence of named entity labels for a given input sentence.

$$P(s|o) = \frac{1}{Z} \exp\left[\sum_{t=1}^{T} \lambda_k f_k(s_{t-1}, s_t, o, t)\right]$$

where Z is the normalization factor obtained by marginalizing over all state sequences, fk (st−1, st, o, t) is an arbitrary feature function and λk is the learned weight for each feature function.By using dynamic programming, state transition between two CRF states can be efficiently calculated. The modified forward values, αT (si), to bethe ”unnormalized probability” of arriving state si given the observations < o1, o2, • • • , ot >. α0(s) is set to probability of starting in each state s, and recursively calculated as :

$\alpha_{t+1}(s) = \Sigma \alpha_t(s') \exp \Sigma \lambda_k f_k(s', s, o, t)$

V. RESULT EVALUATION

Evaluation of Named Entity Recognition (NER) systems designed for newspapers is a critical aspect to ensure their reliability and effectiveness in extracting and categorizing named entities from textual data. To assess the performance of an NER system, several evaluation metrics and methods can be employed.One of the primary metrics used in evaluating an NER system is Precision, which measures the correctness of the extracted named entities. It calculates the ratio of correctly identified entities to the total number of entities predicted by the system. Recall, on the other hand, measures the completeness of the extracted named entities by calculating the ratio of correctly identified entities to the total number of entities present in the dataset. F1 Score, a harmonic mean of precision and recall, provides a balanced measure between these two metrics, offering a comprehensive evaluation of the system's performance.

In addition to these quantitative metrics, several evaluation processes can be adopted to gain deeper insights into the system's effectiveness. Location Evaluation assesses the system's ability to correctly identify and categorize named entities in various text fields such as dates, search terms, and proper names. Error Analysis is another crucial step that involves identifying and categorizing the types of errors made by the NER system. This helps in understanding the system's weaknesses and provides opportunities for improvement.

Cross-Validation is another essential evaluation method where the dataset is split into training and testing sets multiple times. This approach ensures the stability and generality of the system's performance across different data samples. Benchmarking the NER system against existing baselines and state-of-the-art methods provides a comparative analysis, highlighting its relative effectiveness.

Lastly, Human Assessment serves as a complementary, qualitative assessment that captures nuances that may be overlooked by automated metrics. Engaging human evaluators to provide feedback on the system's performance can offer valuable insights into its practical applicability and quality. Considering the diverse nature of textual data in newspapers, it's essential to evaluate the system's performance across various fields to ensure its effectiveness in different contexts.

Testing the NER system on unseen or out-of-sample data can also help assess its ability to generalize and adapt to new information. Ultimately, by employing a comprehensive evaluation process that combines quantitative metrics with qualitative assessments, you can gain a holistic understanding of the NER system's performance, identify areas for improvement, and ensure its reliability in real-world applications.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

VI. CONCLUSION

This report delves into the supervised learning methodologies employed in Named Entity Recognition (NER), a pivotal technique within the realm of Natural Language Processing (NLP). NER focuses on identifying and categorizing entities within unstructured text, enabling machines to comprehend and organize information essential for various applications like text recognition, knowledge mapping, and knowledge graph generation.

- Dataset Preparation and Evaluation Framework

  The initial step in evaluating an NER model involves meticulous dataset partitioning. The dataset is bifurcated into distinct training and test sets. The training set is utilized to train the model, while the test set serves as a rigorous evaluation ground.

  Within this evaluation framework, several key metrics are employed to quantify the model's performance. Precision, Recall, and F1-score stand out as crucial metrics that offer quantitative insights into the model's accuracy, spanning from individual tokens to entire entity sequences.

- Supervised Learning Algorithms in NER

  Supervised learning algorithms form the backbone of NER models. Various algorithms such as Hidden Markov Model (HMM), Decision Trees, Maximum Entropy Model (ME), Support Vector Machine (SVM), and Conditional Random Field (CRF) are employed. Each of these algorithms utilizes training examples to identify patterns and make predictions.

  The primary objective across these algorithms is to optimize model performance by resolving conflicting rules or distributions that maximize the accuracy of entity identification.

- Comparative Analysis and Error Analysis

  To contextualize the performance of an NER model, it is crucial to conduct a comparative analysis against baseline methods.

  This analysis sheds light on potential areas of improvement and innovation. Furthermore, error analysis is instrumental in identifying the root causes of inaccuracies within the model. By pinpointing these sources of errors, targeted refinements can be implemented to enhance the model's efficacy and precision.

- Cross-Validation and Generalization

  Cross-validation techniques are paramount in ensuring the robustness and generalizability of the NER model. By subjecting the model to evaluation across multiple data subsets, its stability and adaptability are validated. Additionally, the model's capability to generalize to unseen data is evaluated by testing it on separate test sets.

  This serves as a critical litmus test for the model's practical deployment and real-world applicability.

VII. REFERENCES

[1] Yoshua Bengio, R´ejean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. J. Mach. Learn. Res., 3:1137–1155, March 2003. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=944919.944966.

[2] Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel. An algorithm that learns whats in a name. Mach. Learn., 34(1-3):211–231, feb 1999. ISSN 0885-6125. doi: 10.1023/A:1007558221122. [4] URL http://dx.doi.org/10.1023/A:1007558221122.

[3] Challagundla, Bhavith Chandra. “Neural Sequence-to-Sequence Modeling with Attention by Leveraging Deep Learning Architectures for Enhanced Contextual Understanding in Abstractive Text Summarization.” International Journal of Machine Learning and Cybernetics (IJMLC), 2024.

[4] Andrew Eliot Borthwick. A maximum entropy approach to named entity recognition. PhD thesis, New York, NY, USA, 1999. AAI9945252.t. In proceedings of the 6th conference on Natural language learning - Volume 20, COLING-02, pages 1– 4, Stroudsburg, PA, USA, 2002.

[5] Association for Computational Linguistics. doi: 10.3115/1118853.1118857. URL http://dx.doi.org/10.3115/1118853.1118857.

[6] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to order things. J. Artif. Int.Res.,10(1): 243–270,May1999.ISSN1076-9757 URL :http://dl.acm.org/citation.cfm?id= 1622859. 1622867. Ronan Collobert and Jason Weston. A unified architecture for natural language processing: deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, ICML '08, pages 160–167, New York, NY, USA, 2008. ACM. ISBN978-1-60558-205-4.doi:10.1145/1390156.1390177.

[7] Bhavith Chandra Challagundla, Chakradhar Reddy Peddavenkatagari, Yugandhar Reddy Gogireddy, "Efficient CAPTCHA Image Recognition Using Convolutional Neural Networks and Long Short-Term Memory", International Journal of Scientific Research in Engineering and Management, Volume 8, Issue 3 DOI : 10.55041/IJSREM29450

[8] Ronan Collobert, Jason Weston, L´eon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. J. Mach. Learn. Res., 12:2493–2537,November2011.ISSN1532-4435.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. In Machine Learning, pages 273–297, 1995.

[10] Challagundla, Bhavith Chandra. "Efficient CAPTCHA Image Recognition Using Convolutional Neural Networks and Long Short-Term Memory Networks." International Journal of Scientific Research in Engineering and Management (IJSREM), 2024. doi:10.55041/IJSREM29450.