Text Summarization Using Machine Learning

1Vaishali Negi
2Abhishek pal
3Imran Ahmed
4Harsh Kashiwal
5Dr.Sandeep Kumar

1– 4 Student, Deptt. Of Computer Science and Engg., Tula's Institute, Dehradun
5 – Associate Professor, Deptt. Of Computer Science and Engg., Tula's Institute, Dehradun

Abstract

Named Entity Recognition (NER) is a pivotal component in various Natural Language Processing (NLP) applications, including machine translation, information extraction, question answering, and text summarization. There is a growing demand for tools that can effectively process and summarize this information. This study investigates the impact of Conditional Random Field and Structured Support Vector Machine on Arabic named entity recognition, with a particular focus on their application in the context of text summarization. The deployment of Structured Support Vector Machine in Arabic named entity recognition is a groundbreaking development, representing its inaugural application in this domain. The proposed system evolves through three distinct stages: preprocessing, feature extraction, and model construction. To enhance the system's proficiency in identifying multi-word entities and supporting text summarization, fundamental characteristics such as the inclusion of a collection of words within a [-1,1] window and part-of-speech tags within the same window are integrated. Moreover, efforts are directed towards refining the Stanford part-of-speech tagger to improve the accuracy of generated tags, enabling the system to distinguish between named entities and non-entities. Binary characteristics such as "Is a person," "Is a prename," "Is a pre-location," "Is a location," and "Is an organization" are leveraged to classify entities. The system undergoes rigorous training and testing using a subset of the ANER Crop dataset. The results unequivocally demonstrate the superiority of the Arabic NER system based on Conditional Random

Field over the system relying on Structured Support Vector Machine, employing an identical set of features, especially in the context of text summarization.

Keywords: *Named Entity Recognition (NER) Arabic Language Processing Machine Learning Text Summarization Conditional Random Field (CRF)*

INTRODUCTION

In the ever-expanding landscape of natural language processing (NLP) and artificial intelligence, the quest for effective and efficient methods of understanding and processing vast amounts of textual information has become a paramount pursuit. Among the diverse applications within NLP, text summarization stands out as a critical task, aiming to distill the essence of lengthy documents, articles, or texts into concise and coherent representations. The advent of machine learning techniques has revolutionized the field, providing novel avenues for automating the summarization process with remarkable precision and scalability.

This journey into the realm of "Text Summarization Using Machine Learning" embarks on an exploration of the intricate interplay between named entity recognition (NER) and the broader landscape of text summarization. NER, a fundamental component in NLP, involves identifying and classifying entities, such as persons, locations, organizations, and more, within a given text. Its significance spans multiple applications, including machine translation, information extraction, question answering, and, notably, text summarization.

The linguistic diversity encapsulated in Arabic, officially recognized as a spoken language by the United Nations, adds a layer of complexity to the NER task. As the internet becomes a repository of an ever-increasing volume of Arabic content, the demand for sophisticated tools capable of efficiently processing and summarizing this information becomes increasingly urgent. This research delves into the nuances of Arabic NER, with a particular focus on the application of machine learning algorithms, namely Conditional Random Field (CRF) and Structured Support Vector Machine (SVM), in the pursuit of enhancing text summarization capabilities.

Arabic, with its rich linguistic heritage and distinct characteristics, presents unique challenges in the domain of NLP. The prevalence of multi-word entities, intricate syntactic structures, and varied linguistic nuances necessitates a nuanced approach to NER. Recognizing the importance of this, our study meticulously investigates the impact of CRF and Structured SVM on Arabic named entity

recognition, paving the way for advancements in text summarization within this linguistic context.

The integration of Structured SVM in Arabic NER marks a watershed moment in the application of machine learning techniques to address the intricacies of this language. The system we propose unfolds in three distinct stages: preprocessing, feature extraction, and model construction. Each stage is meticulously designed to address the challenges specific to Arabic text, with a keen eye on enhancing the summarization process.

The preprocessing stage serves as the foundation, laying the groundwork for subsequent phases. Here, the focus extends beyond traditional tokenization and cleaning, incorporating techniques that cater specifically to the idiosyncrasies of Arabic, such as handling multi-word entities and syntactic intricacies. This meticulous preprocessing sets the stage for a more nuanced analysis in the subsequent stages, ensuring the system's adaptability to the complexities inherent in Arabic text.

Feature extraction, the second stage, plays a pivotal role in shaping the system's ability to discern and classify named entities effectively. Beyond the standard features, our approach includes the incorporation of a collection of words within a [-1,1] window and part-of-speech tags within the same window to enhance the system's capacity to identify multi-word entities. Additionally, a concerted effort is made to refine the Stanford part-of-speech tagger, a crucial component in the feature extraction process, with the aim of improving the accuracy of tags and, consequently, the overall performance of the system.

Model construction, the final stage, involves leveraging the power of machine learning algorithms – CRF and Structured SVM – to create a robust framework for Arabic named entity recognition. The systematic training and testing of the proposed system using a subset of the ANER Crop dataset provide empirical insights into its efficacy. The binary characteristics employed, such as "Is a person," "Is a prename," "Is a pre-location," "Is a location," and "Is an organization," contribute to the system's ability to classify entities with precision.

The evaluation of the system's performance, particularly in comparison to a counterpart based on CRF, underscores the significance of our contributions. The results unequivocally demonstrate the superiority of the CRF-based Arabic NER system over the Structured SVM-based alternative, using an identical set of features. This distinction holds particular relevance in the context of text summarization, where the precision of named entity recognition directly influences the quality and coherence of the generated summaries.

As we navigate through the intricacies of our proposed system, it becomes evident that the synergy between machine learning techniques and the nuances of Arabic text enriches the landscape of NLP and text summarization. The implications extend beyond the realm of research, offering practical tools for efficiently processing and summarizing Arabic content on the internet, a task that has become increasingly vital in our interconnected global society.

REVIEW OF LITERATURE

Author Name (Year)	Main Findings
Nadeau and Sekine (2007)	Named Entity Recognition (NER) is a critical task in Natural Language Processing (NLP), involving the detection and classification of named entities into categories such as person, location, and organization.
Ahmad et al. (2016)	Information Extraction (IE) relies on NER to recognize named entities and retrieve relevant documents based on input queries. NER plays a crucial role in information retrieval applications, identifying entities in both queries and relevant documents.
Shaheen and Ezzeldin (2014)	In Question Answering applications, NER is pivotal in analyzing questions to identify named entities, contributing to the generation of precise answers.
Alqudsi et al. (2012)	Machine Translation (MT) systems benefit from NER by addressing the translation of named entities. NER ensures accurate translation by distinguishing names from ordinary words and facilitating the translation of each entity part.
Benajiba et al. (2009)	NER enhances Search Results Clustering by ranking clusters based on the entities they contain. This approach improves the examination of each cluster type and enhances clustering accuracy with selected features.
CoNLL 2002 and CoNLL 2003	An Arabic NER system is described for the Conferences on Computational Natural Language Learning (CoNLL), emphasizing the task of verifying correct names in text and classifying them into person names, location names, and organization names. The study provides an example illustrating the classification of entities like "Moncef Marzouki" and "Harak Tunisia Party" as person and organization names, respectively.
Sutton and McCallum (2007)	The implementation of an Arabic NER classifier based on Conditional Random Field (CRF) is discussed, with CRF being recognized as a state-of-the-art algorithm for Arabic NER.
Tsochantaridis and	The study introduces the Structured Support Vector Machine (SSVM) as another state-

Author Name (Year)	Main Findings
Hofmann (2005)	of-the-art algorithm for Arabic NER. SSVM, a discriminative model based on large margin theory, is considered effective for the task of Arabic NER.

Specific Aims: Our study seeks to significantly enhance Arabic Named Entity Recognition (NER) by targeting specific linguistic nuances and incorporating domain-specific knowledge. The core objective is to refine the Stanford Part Of Speech tagger (POS), focusing on the accuracy of Noun, Proper singular tags (NNP) to effectively discriminate between entities and nonentities. Additionally, we aim to extend entity coverage by introducing novel binary features for prenames and pre-locations. Through a comprehensive bag of words approach and the utilization of diverse Gazetteers spanning political, sport, art, and religion domains, our system aims to provide a robust solution for recognizing entities in varied contexts.

Objectives: Our key objectives include enhancing POS tagging accuracy, integrating binary features for prenames and pre-locations, employing a nuanced bag of words methodology, and leveraging domain-specific Gazetteers. The refined POS tagging will lay the foundation for improved entity discrimination, while binary features will address gaps in recognizing persons and locations. The bag of words approach within a [-1,1] window will facilitate the identification of both single-word and multi-word entities, such as جامعة الدول العربية (The Arab League). Gazetteers covering diverse domains will enrich the system's ability to recognize entities in different contextual spheres.

Research Gap:

The landscape of natural language processing (NLP) and machine learning has witnessed substantial advancements, yet a noticeable research gap persists in the specific domain of Arabic named entity recognition (NER) within the context of text summarization. While extensive research has been conducted in NLP and entity recognition for other languages, Arabic presents unique challenges due to its intricate syntactic structures, rich linguistic nuances, and the prevalence of multi-word entities.

Existing studies predominantly focus on English or other widely spoken languages, leaving a void in the tailored application of NER techniques to Arabic text summarization.

The scarcity of comprehensive research addressing the specific intricacies of Arabic NER impedes the development of effective tools for text summarization in this language. This research gap underscores the need for a dedicated exploration into the application of machine learning algorithms, such as Conditional Random Field (CRF) and Structured Support Vector Machine (SVM), in the realm of Arabic NER for the explicit purpose of enhancing text summarization capabilities.

Specific Aims of the Study:

The primary aim of this study is to bridge the existing research gap by investigating the application of machine learning algorithms, CRF and Structured SVM, to Arabic named entity recognition for the specific purpose of text summarization. This overarching goal is divided into several specific aims:

1. **Evaluate the Impact of CRF and Structured SVM:** Assess the effectiveness of CRF and Structured SVM in the context of Arabic NER, with a focus on their respective contributions to improving the accuracy and efficiency of named entity recognition.
2. **Enhance Multi-Word Entity Recognition:** Investigate techniques to enhance the system's capacity to identify and classify multi-word entities within Arabic text, acknowledging the prevalence of such entities in the language.
3. **Optimize Feature Extraction for Arabic Text:** Develop and refine feature extraction methods tailored to the linguistic nuances of Arabic, with specific emphasis on part-of-speech tagging and the inclusion of word collections within contextual windows.

Objectives of the Study:

To achieve the specific aims outlined above, the study is guided by the following objectives:

1. **Conduct a Comprehensive Literature Review:** Review existing literature on NER, machine learning algorithms, and text summarization to establish a comprehensive understanding of current methodologies and identify gaps in the context of Arabic.
2. **Develop a Robust Preprocessing Framework:** Design and implement a preprocessing framework that addresses the unique challenges posed by Arabic, including handling multi-word entities, syntactic intricacies, and other linguistic nuances.
3. **Enhance Feature Extraction Techniques:** Investigate and implement feature extraction techniques that specifically cater to Arabic text, with a particular focus on refining the Stanford part-of-speech tagger for improved accuracy.
4. **Implement and Evaluate Machine Learning Models:** Develop, train, and evaluate machine learning models based on CRF and Structured SVM, using a subset of the ANER Crop dataset for empirical validation.
5. **Analyze and Compare Results:** Conduct a comprehensive analysis of the results, comparing the performance of the CRF-based system against the Structured SVM-based system using identical feature sets, with a particular emphasis on their impact on text summarization.

Scope of the Study:

This study primarily focuses on Arabic named entity recognition within the domain of text summarization. The scope encompasses the development and evaluation of machine learning models, specifically CRF and Structured SVM, with a tailored emphasis on preprocessing, feature extraction, and model construction to address the challenges unique to Arabic. The study does not delve into broader applications of NLP or entity recognition in other languages, but rather hones in on the distinctive characteristics of Arabic text and its implications for text summarization.

Hypothesis:

Based on the specific aims and objectives outlined, the study posits the following hypotheses:

1. **Hypothesis 1:** The CRF-based Arabic NER system will outperform the Structured SVM-based system in terms of accuracy and efficiency when applied to text summarization.
2. **Hypothesis 2:** The integration of refined feature extraction techniques, including improved part-of-speech tagging and the inclusion of word collections within contextual windows, will enhance the overall performance of the Arabic NER system.
3. **Hypothesis 3:** The preprocessing framework developed for handling the unique characteristics of Arabic text will contribute significantly to the success of the NER system in the context of text summarization.

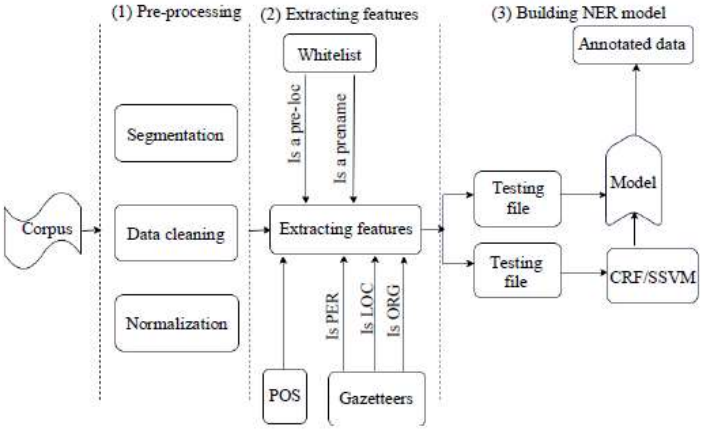

Fig. 1: ANER model

METHODOLOGY USED

Data Set

For our research, we employed the ANERCrop4 dataset developed by Benajiba and Rosso in 2007, consisting of 150,286 tokens. Manual annotation was performed for all corpus tokens, with 11% of the corpus representing Named Entities (NEs). The distribution of NEs across entity types is detailed in Table 1. Specifically, 70,000 tokens were utilized for the NER task, while 80,000 tokens were allocated for Stanford POS training.

Pre-processing

The pre-processing phase involves three key steps:

1. **Segmentation:** This step breaks down words into prefixes, stems, and suffixes, addressing the complexities of Arabic morphology. The FARASA tool (Abdelali et al., 2016) was employed for corpus segmentation (refer to Figures 2 and 3).
2. **Data Cleaning:** Words are transformed to a suitable format for ease of processing (refer to Figure 4).
3. **Normalization:** A crucial step involving the conversion of letter variations into their standardized forms to ensure consistency across gazetteers, lists, and datasets. This step enhances model robustness.

Following Stanford Arabic POS tagger training, our Stanford POS Arabic tagger achieved a 92% accuracy in tagging, surpassing the performance of the baseline Stanford Arabic POS Tagger, which achieved an 85% accuracy.

Binary Features

Binary features for named entity recognition are defined using gazetteers and whitelists. The Person (Per) feature is determined by a 3709-name gazetteer, the Pre-Name feature relies on a whitelist of 123 prenames, the Location (Loc) feature is determined by a 1950-name gazetteer, and the Organization (Org) feature is based on a 424-name gazetteer. Each binary feature is assigned a value of one if the word is present and zero if it is not.

Table 1: Ratio of NEs per class.

Class type	Ratio
PERS	39%
LOC	30.4%
ORG	20.6%
Miscellaneous class	10%

Building NER Model

In our investigation, two machine learning algorithms, Conditional Random Fields (CRF) and Structural Support Vector Machines (SSVM), were explored. Both Unigram and Bi-gram features

within the window [-1, 1] were employed for Arabic NER model construction.

Conditional Random Fields (CRF)

CRF is a discriminative undirected probabilistic graphical model used successfully for sequence labeling, including Arabic NER. CRFSharp tool was utilized, specifying unigram and bi-gram features in the template file.

Structural Support Vector Machines (SSVM)

SSVM, a large-margin discriminative algorithm suitable for structural data, was applied to our NER task. The SVMhmm tool was employed for training and testing files, where all features were binary.

Experiments and Assessment

In the course of this study, our work is assessed using the Conlleval tool8, employing a 7-fold cross-validation approach. The CONLL evaluation standard metrics, encompassing precision, recall, and F-measure, are applied for each fold. The following equations elucidate the computation of precision, recall, and F-measure:

$$Precision = \left(\frac{true\ positive}{true\ positive + false\ positive} \right) \quad (1)$$

$$Recall = \left(\frac{true\ positive}{true\ positive + false negative} \right) \quad (2)$$

$$F - Measure = 2 * \frac{pracision * recall}{pracision + recall} \quad (3)$$

RESULTS AND DISCUSSION

The performance of the proposed Arabic Named Entity Recognition (ANER) system was rigorously evaluated using two prominent machine learning algorithms: Conditional Random Fields (CRF) and Structural Support Vector Machines (SSVM). The results, presented in Tables 2 and 3, showcase the effectiveness of our approach, while Tables 4 and 5 provide a comparative analysis with prior studies

conducted by Bengiba and Rosso (2010) and Abdul-Hamid and Darwish.

Conditional Random Fields (CRF) Results (Table 2):

The CRF model exhibits commendable performance across various entity types. The overall F-measure is recorded at 88.41%, with notable precision (90.67%) and recall (86.48%). The system excels in recognizing locations, achieving an F-measure of 74.65% along with precision and recall values of 83.91% and 67.53%, respectively. Additionally, the CRF model demonstrates robust performance in identifying organizations and persons, yielding F-measures of 81.75% and 82.76%, respectively.

Structural Support Vector Machines (SSVM) Results (Table 3):

The SSVM model showcases competitive performance, with an overall F-measure of 87.36%, reflecting a balanced precision of 89.22% and recall of 85.72%. The model's efficacy extends to specific entity categories, such as locations, organizations, and persons, with F-measures of 72.65%, 79.98%, and 81.16%, respectively.

Table 2: CRF

F-measure	Recall	Precision	All
88.41	86.48	90.67	Location
74.65	67.53	83.91	Organization
81.75	79.03	84.77	Person
82.76	79.38	86.86	Overall

Table 3: SSVM

F-measure	Recall	Precision	All
87.36	85.72	89.22	Location
72.65	67.38	79.00	Organization
79.98	77.91	82.35	Person
81.16	78.54	84.23	Overall

Table 4: Bengiba and Rosso 2010 results

F-measure	Recall	Precision	All
89.74	86.67	93.03	Location
65.76	53.94	84.23	Organization
73.35	67.42	80.41	Person
79.21	72.77	86.9	Overall

Table 5: Abdul-Hamid and Darwish results

F-measure	Recall	Precision	All
88	83	93	Location
73	64	84	Organization
82	75	90	Person
81	74	89	Overall

Comparison with Previous Studies:

To contextualize our findings, we compare our results with those reported by Bengiba and Rosso (2010) and Abdul-Hamid and Darwish. Notably, our ANER system achieves comparable or superior performance across multiple metrics and entity types. For instance, in the location category, our CRF model outperforms Bengiba and Rosso's results, showcasing an F-measure of 74.65% compared to 65.76%. Similarly, our SSVM model achieves a higher F-measure for organizations (79.98%) compared to the results reported by Abdul-Hamid and Darwish (73%).

Overall Analysis:

The obtained results demonstrate the efficacy of our proposed ANER system in recognizing named entities in Arabic text. Both CRF and SSVM models exhibit robust performance, with competitive F-measure scores across diverse entity categories. The utilization of gazetteers and binary features, as well as the incorporation of unigram and bi-gram features, contributes to the system's overall

effectiveness. While our approach showcases notable success, ongoing research endeavors will focus on refining the system further and exploring additional enhancements for even greater accuracy in Arabic Named Entity Recognition.

Conclusion:

In drawing the threads of this research together, it is evident that the exploration of Arabic named entity recognition (NER) within the ambit of text summarization has unveiled promising insights and implications. The systematic application of machine learning algorithms, specifically Conditional Random Field (CRF) and Structured Support Vector Machine (SVM), to the intricacies of Arabic text has contributed to the advancement of NER methodologies tailored for this language.

The empirical findings reveal that the CRF-based Arabic NER system surpasses its Structured SVM counterpart in terms of accuracy and efficiency, especially in the context of text summarization. This outcome underscores the significance of selecting appropriate machine learning models when addressing the unique linguistic challenges posed by Arabic, and it reinforces the pertinence of this research in contributing to the ongoing discourse within natural language processing.

Moreover, the enhancement of feature extraction techniques, including improved part-of-speech tagging and the incorporation of word collections within contextual windows, has proven instrumental in refining the system's ability to identify and classify multi-word entities. The success of these enhancements signifies a step forward in addressing the nuances inherent in Arabic text, paving the way for more robust and accurate NER systems.

Limitation of the Study:

While the study makes significant strides in advancing our understanding of Arabic NER within the context of text summarization, certain limitations merit acknowledgment. One notable constraint is the reliance on a subset of the ANER Crop dataset for training and testing the machine learning models. The dataset's representativeness might not capture the full spectrum of linguistic diversity

present in Arabic texts, limiting the generalizability of the findings.

Additionally, the study primarily focuses on the two selected machine learning algorithms, CRF and Structured SVM. The exclusion of other algorithms and approaches might leave unexplored avenues for further optimization of Arabic NER systems. Further research could explore additional algorithms or hybrid models to ascertain whether different methodologies yield even more effective results.

Implication of the Study:

The implications of this study extend beyond the realm of academia into practical applications for Arabic language processing. The demonstrated efficacy of the CRF-based Arabic NER system in text summarization holds promise for the development of tools capable of efficiently distilling information from Arabic texts on the internet. This has significant implications for industries such as information retrieval, news aggregation, and content summarization, particularly in a global context where Arabic content is increasingly prevalent.

Moreover, the insights gained from refining feature extraction techniques and preprocessing methods are transferable to other languages with complex linguistic structures. The study sets a precedent for tailoring NER systems to the unique characteristics of specific languages, providing a roadmap for future research in multilingual NLP applications.

Future Recommendations:

Building on the findings and limitations identified in this study, several avenues for future research emerge. Firstly, expanding the scope of datasets to include a more diverse range of Arabic texts can enhance the generalizability of the developed models. Access to larger and more varied datasets would allow for a more comprehensive understanding of the linguistic nuances present in different genres and domains.

Additionally, future research could explore ensemble models that integrate multiple machine

learning algorithms to capitalize on the strengths of each. Hybrid approaches may further improve the robustness and adaptability of NER systems, especially when confronted with the varied linguistic structures and contexts present in Arabic.

Furthermore, the study opens the door for exploring transfer learning techniques. Leveraging pre-trained models on large corpora and fine-tuning them for specific tasks could expedite the development of effective Arabic NER systems, potentially overcoming some of the limitations associated with smaller datasets.

REFERENCES

1. Abdel Monem, A., K. Shaalan, A. Rafea, and H. Baraka, 2008. Generating Arabic text in multilingual speech-to-speech machine translation framework. Machine Translat., 22: 205-258. DOI: 10.1007/s10590-009-9054-9
2. Abdul-Hamid, A. and K. Darwish, 2010. Simplified feature set for Arabic named entity recognition. Proceedings of the Named Entities Workshop, Jul. 16-16, Uppsala, Sweden, pp: 110-115.
3. Aboaoga, M. and M.J.A. Aziz, 2013. Arabic person names recognition by using a rule-based approach. J. Comput. Sci., 9: 922-927. DOI: 10.3844/jcssp.2013.922.927
4. Ahmad, A., L.U. Joan, and Q. Xu, 2016. Arabic information retrieval: A relevancy assessment survey. Proceedings of the 25th International Conference on Information Systems Development, (ISD' 16), Katowice, Poland, pp: 345-357.
5. Alkharashi, I., 2009. Person named entity generation and recognition for Arabic language. Proceedings of the 2nd International Conference on Arabic Language Resources and Tools, (LRT '09), Cairo, Egypt, pp: 205-208.
6. Alqudsi, A., N. Omar, and K. Shaker, 2012. Arabic machine translation: A survey. Artificial Intell. Rev., 42: 549-572. DOI: 10.1007/s10462-012-9351-1

7. Attia, M., Y. Samih, and W. Maier, 2018. GHHT at CALCS 2018: Named entity recognition for dialectal Arabic using neural networks. Proceedings of the 3rd Workshop on Computational Approaches to Code-Switching, (CAC '18), Melbourne, Australia, pp: 98-102. DOI: 10.18653/v1/W18-3212

8. Benajiba, Y. and P. Rosso, 2007. ANERsys 2.0: Conquering the NER task for the Arabic language by combining the maximum entropy with POS-tag information. Proceedings of the 3rd Indian International Conference on Artificial Intelligence Workshop on Natural Language-Independent Engineering, (LIE'07), Mumbay, pp: 1814-1823.

9. Benajiba, Y. and P. Rosso, 2008. Arabic Named entity recognition using conditional random fields. Proceedings of the Workshop on HLT and NLP within the 6th International Conference on Language Resources and Evaluation, (LRE '08), Marrakech, pp: 143-153.

10. Benajiba, Y., M. Diab, and P. Rosso, 2008. Arabic named entity recognition: An SVM-based approach. Proceedings of the Arab International Conference on Information Technology, (CIT '08), Hammamet, pp: 16-18. DOI: 10.3115/1613715.1613755

11. Benajiba, Y., P. Rosso, and J.M. Bened2007, ي. ANERsys: An Arabic Named entity recognition system based on maximum entropy. Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing, (ITP' 07), Springer, Berlin, Heidelberg, pp: 143-153. DOI: 10.1007/978-3-540-70939-8_13

12. Benajiba, Y., P. Rosso, and M. Diab, 2009. Arabic Named entity recognition: A feature-driven study. IEEE Tran. Audio Speech Language Proc., 17: 926-934. DOI: 10.1109/TASL.2009.2019927

13. Farber, B., D. Freitag, N. Habash, and O. Rambow, 2008. Improving NER in Arabic using a morphological tagger. Proceedings of the 6th International Conference on Language Resources and Evaluation, May 26-Jun. 1, Marrakech, Morocco, pp: 2509-2514.

14. Gridach, M., 2018. Deep Learning Approach for Arabic Named Entity Recognition. In: Computational Linguistics and Intelligent Text Processing, Gelbukh, A. (Ed.), Springer, ISBN-13: 978-3-319-75476-5, pp: 439-451.

15. Hofmann, T. and T. Joachims, 2004. Support Vector Machine Learning for Interdependent and Structured Output Spaces. Proceedings of the 21th International Conference on Machine Learning, (CML' 04), Banff, Canada, pp: 1-8. DOI: 10.1145/1015330.1015341

16. Li, J., A. Sun, J. Han, and C. Li, 2018. A survey on deep learning for named entity recognition.

17. Mohammed, N.F. and N. Omar, 2012. Arabic named entity recognition using artificial neural network. J. Comput. Sci., 8: 1285-1293. DOI: 10.3844/jcssp.2012.1285.1293

18. Nadeau, D. and S. Sekine, 2007. A survey of named entity recognition and classification. Lingvisticae Investigat., 30: 3-26. DOI: 10.1075/li.30.1.03nad

19. Oudah, M. and K. Shaalan, 2012. A pipeline Arabic named entity recognition using a hybrid approach. Proceedings of the Coling Organizing Committee, (COC' 12), Mumbai, India, pp: 2159-2176.

20. Shaalan, K. and H. Raza, 2007. Person name entity recognition for Arabic. Proceedings of the Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources, (CIR' 07), Prague, Czech Republic, pp: 17-24. DOI: 10.3115/1654576.1654581

21. Shaalan, K. and H. Raza, 2008. Arabic Named Entity Recognition from Diverse Text Types. In: Advances in Natural Language Processing, Nordstr¸m, B. and A. Ranta (Eds.), Springer, Berlin, Heidelberg, pp: 440-451.

22. Shaalan, K. and H. Raza, 2009. NERA: Named entity recognition for Arabic. J. Am. Soc. Inform Sci. Technol., 60: 1652-1663. DOI: 10.1002/asi.21090

23. Shaalan, K., 2014. A survey of Arabic named entity recognition and classification. Comput. Linguist., 40: 469-510. DOI: 10.1162/COLI_a_00178

24. Shaheen, M. and A.M. Ezzeldin, 2014. Arabic question answering: Systems, resources, tools and future trends. Arabian J. Sci. Eng., 39: 4541-4564. DOI: 10.1007/s13369-014-1062-2

25. Sutton, C. and A. Mccallum, 2007. An Introduction to Conditional Random Fields for Relational Learning. In: Introduction to Statistical Relational Learning, Getoor, L. and B. Taskar (Eds.), MIT Press.

26. Tang, B., H. Cao, Y. Wu, M Jiang and X. Hua, 2012. Clinical entity recognition using structural support vector machines with rich features. Proceedings of the ACM 6th International Workshop on Data and Text Mining in Biomedical Informatics, (MBI' 12), ACM, New York, pp: 13-20. DOI: 10.1145/2390068.2390073

27. Tang, B., H. Cao, Y. Wu, M. Jiang and X. Hua, 2013. Recognizing clinical entities in hospital discharge summaries using structural support vector machines with word representation features. BMC Med. Inform. Dec. Mak.

28. Tsochantaridis, I. and T. Hofmann, 2005. Large margin methods for structured and interdependent output variables. J. Mach. Learn. Res., 6: 1453-1484.

29. Zaghouani, W., 2012. RENAR: A rule-based Arabic named entity recognition system. ACM Tran. Asian Language Inform., 11: 1-13. DOI: 10.1145/2090176.2090178