JOURNAL OF APPLIED INTELLIGENT SYSTEMS & INFORMATION SCIENCES
VOL. 2, ISSUE 1, PP. 41-48, JUNE 2021.
Available at: **www.journal.research.fanap.com**
DOI: https://doi.org/10.22034/jaisis.2021.306194.1034

ML REVOLUTION IN NLP: A REVIEW OF MACHINE LEARNING TECHNIQUES IN NATURAL LANGUAGE PROCESSING

Majid Abedinzadeh Shahri [1,*], Rasool Taban [2]

[1] *School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran*
[2] *Center for Computational and Stochastic Mathematics, University of Lisbon, Lisbon, Portugal*

ABSTRACT

In the current era, usually, people communicate with each other through the internet. This platform opens this opportunity for computer scientists to access huge information about human languages. However, this big data is unstructured from the computational point of view. Thus, computer scientists developed Natural Language Processing (NLP) methods for analyzing human language data by computers. Indeed, NLP is a way to analyze human language messages by computerized methods. On the other hand, due to the high capabilities of Machine Learning (ML), many researchers incorporated this approach in language processing techniques to improve the performance of NLP systems. In this paper, we aim to present a summarized review of the NLP techniques. considering the importance of the ML approach. Firstly, this paper introduces basic terminology for NLP. Secondly, according to the importance of ML history, the studied techniques are categorized into three groups: old-fashioned, conventional, and modern methods. The presented review in this study could be beneficial for ML and NLP reseachers in order to develop new ML techniques for NLP tasks.

KEYWORDS: Human language data, Natural Language Processing, Machine Learning

1. INTRODUCTION

Advancements in computer technologies yield this opportunity to store and process big data. On the other hand, with the development of the Internet, access to large amounts of data hase become possible. Additionally, the Internet provides new content-sharing services that allow people to create and share their content in an efficient way with millions of people. However, this abundance of information is transferred via human languages that are unstructured from the computational point of view. To tackle this problem, the researchers tried to propose several computational methods to automatically analyze human language data.

In text analysis, though simple algorithms (e.g., spell-checking) are working well in word-level processing, these methods have no potential to interpret sentences. Hence, NLP requires more capabilities to handle natural language texts. Here, it should be emphasized that in this context, the "natural" term (in NLP) does not refer to formal languages such as mathematical notations (Jackson & Moulinier, 2007)

In (Liddy, 2001), NLP was defined as a collection of computational techniques for analyzing and representing the contents of the human language to achieve human-like language processing. Recently there have been four key factors affecting the noticeable advancement in NLP: a) the remarkable growth of computing

[*] *Corresponding Author, email: m.abedinzadeh@ut.ac.ir*

RECEIVED: 11 MARCH, 2021; **ACCEPTED**: 18 JUNE, 2021; **PUBLISHED ONLINE**: 25 JUNE, 2021

power, b) the enormous amount of electronic texts including published documents such as electronic dictionaries, c) encyclopedias, and libraries (Bose, 2004), d) the advancement in the highly successful ML methods and the deeper understanding of the structure of human language and its deployment in social contexts. Accordingly, several computational tools were developed for NLP (Jusoh & Al-Fawareh, 2007; Kaur et al., 2014; Vijayarani et al., 2015) in different disciplines such as computer and information sciences, linguistics, mathematics, electrical and electronic engineering, psychology, cognitive science, artificial intelligence, robotics, etc.

Over time, NLP is developing important technologies to create user-friendly tools in many aspects of life (Church & Rau, 1995). Machine translation, perhaps as the oldest application of NLP, tries to translate text (or speech) content from one language to another one. Information extraction, as another popular NLP task, is used for automatically extracting structured information from unstructured and/or semi-structured in natural language texts. Named-Entity Recognition is a sub-task of information extraction. Another NLP application is automated text summarization. This is the process of generating a short, readable, and meaningful summary from a long text document (Mani & Maybury, 2001) that may not convey all the details of the actual text. Indeed, it aims to present the basic idea of the original text (Maybury, 1999). Text categorization (Devasena & Hemalatha, 2012; Sebastiani, 2002) is another application of NLP. This is the task of assigning predefined categories to a text document. It can provide conceptual views of document collections and has several applications (such as spam filtering). Question and answering, as another application of NLP, is a system that can automatically understand and build answers to user questions that are posed in a natural language (Al-Harbi et al., 2017; Hirschman & Gaizauskas, 2001).

Nevertheless, in these popular NLP tasks, understanding the meaning of the human message is very difficult due to four reasons: First, a message in a specific language may contain a lot of coded symbols. Second, some of the words in the language have multiple meanings. Also, some words behave differently when used as nouns or verbs. Third, different human languages are structured differently by specific grammatical rules. Fourth, in general, human language texts usually contain ambiguity. The ambiguity is the result of the semantic gap between the user's intentions and how the message is conveyed. Hence, to understand a human language, various powerful analysis techniques (such as ML) have been considered (Agarwal, 2017).

Most ML techniques employ training data in order to optimize performance criteria. Accordingly, for these methods, a model is defined based on some parameters, and learning which executes a computer program uses the training data to optimize the parameters of the model. The model might be predictive or descriptive or both, to make predictions for the future or obtain knowledge from data. Recently, computer vision, speech recognition, and robotics have also been using ML for their solutions. In this era where machines are becoming more intelligent, according to the importance of ML, a lot of NLP research works, which used ML techniques, have been published so far. Considering the ML history, this paper aims to categorize NLP research into three groups: old-fashioned, conventional, and modern methods. It is worth noting that covering different ML methods is regarded as a criterion to choose papers for review.

The rest of the paper is structured as follows; Section 2 presents the basic terminology of NLP. Then, several NLP approaches presented in the literature are studied in Section 3. Finally, the paper ends with discussions and conclusions in Section 4.

2. NLP & Language Understanding

For every NLP task, natural language understanding is an important issue. The meaning of a message is formed in the language by the relations and differences between its shorter components. Accordingly, an NLP system should begin at the word level to determine the nature of each word (known as the Part of Speech tagging), and then, it can move on to the sentence level to determine the meaning of the entire sentence; and finally, to the context and the overall domain. Hence, an NLP system to understand a natural language should be able to distinguish among the following seven levels of processing (Liddy, 1998):

- *Phonology*: This level which includes three types of rules involves the analysis of speech sounds. The rules are 1) the rules for sounds within words, 2) the rules for variations of pronunciation when words are spoken together, and, 3) the rules for fluctuation in stress and intonation across a sentence. The sound waves are analyzed and encoded into a digitized signal to be interpreted using various methods in an NLP system that accepts spoken input.

- *Morphology*: The componential nature of words, which are composed of morphemes (the smallest units of meaning), are analyzed in this technique. A new word is broken down into its morphemes by the human to understand its meaning, so similarly, an NLP system can recognize the meaning of a conveyed word.
- *Lexical*: This level deals with defining the meaning of each word by an NLP system. Words with only one possible meaning can be replaced by a semantic representation of that meaning. Based on the semantic theory used in the NLP system, the nature of the representation differs.
- *Syntactic*: Here the words of a sentence are analyzed to uncover the grammatical structure of a sentence. In fact, the structural dependency relationships between the words are revealed by these analyses which outcomes a representation of the sentence. In many languages, syntax conveys meaning because order and dependency contribute to meaning. For example, there is a difference between the two sentences: “The cat chased the mouse.” and “The mouse chased the cat.” only from a syntax perspective, but still two different meanings are presented by them.
- *Semantic*: Though most people think semantic processing determines the meaning of a text, it is all the levels that contribute to the meaning. In fact, the possible meanings of a sentence are determined by semantic processing and this is achieved by focusing on the interactions among word-level meanings in the sentence. This level of processing can include the semantic disambiguation of words with multiple senses; in an analogous way to how syntactic disambiguation of words that can function as multiple Parts-Of-Speech is accomplished at the syntactic level. One sense of polysemous words can be selected by semantic disambiguation. The semantic, not the lexical level, would do the disambiguation, in case more information from the rest of the sentence was required for the disambiguation.
- *Discourse*: The discourse level of NLP works with units of text longer than a sentence while syntax and semantics work with sentence-length units. It should be mentioned that these techniques do not interpret multi-sentence texts as just chained sentences. Instead of interpreting the text sentence by sentence at this level, discourse focuses on the properties of the text as a whole that brings meaning by making connections between component sentences.
- *Pragmatic*: Here, the purposeful use of language in situations is dealt with and the context over and above the contents of the text is utilized for understanding. This level is aimed at explaining how extra meaning is read into texts without actually being encoded in them. To achieve this, we need a

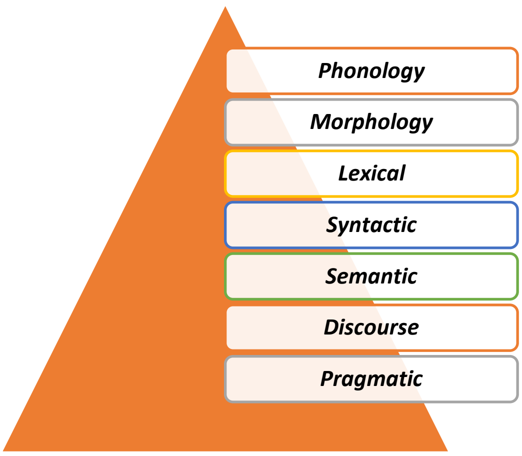

Fig. 1. The seven layers of an NLP system for natural language understanding

lot of world knowledge, including the understanding of intentions, plans, and goals. Knowledge bases and inferencing modules are utilized in some NLP models. For instance, the following two sentences require resolution of the anaphoric term ‘they’, but this resolution requires pragmatic or world knowledge: “The city councilors refused the demonstrators a permit because they feared violence.” and “The city councilors refused the demonstrators a permit because they advocated revolution.”.

According to the aforementioned levels of NLP (see Fig. 1), researchers tend to implement modules for lower levels and combine the implemented lower-level techniques to develop modules for higher levels. This is because of the fact that the application may not require processing at the higher levels. Additionally, the lower levels deal with smaller units (e.g., morphemes, words, and sentences), versus higher levels deal with overall texts.

3. NLP Techniques

In the following, the studied NLP works are categorized into three ages.

3.1. *NLP before ML*

Since syntactic processing was essential in most language processing applications, in the early days, most NLP researches have been focusing on syntax-based processing. The article entitled “Computing Machinery and Intelligence” is one of the earliest works in NLP (Turing & Haugeland, 1950). Artificial intelligence and linguistics methods are combined in this work. Nevertheless, the lexical ambiguity of natural language was ignored by the systems made in the early years. Consequently, this approach resulted in poor outcomes. The need for external knowledge in interpreting and responding to language input was recognized in the later work.

The concept of production rule was proposed in another work in this field (Chomsky, 1956) to describe natural language. A production rule includes a set of conditions and a consequent set of actions. ‘Recognize’, ‘resolve conflict’, and ‘act’ that is repeated until no more rules have remained are the main actions of a production rule system. The rules are independent of each other and this allows each rule to be added and deleted easily. Human also understands these rules easily because the rules are usually derived from the observation of expert knowledge. However, there are scalability issues when production rule systems become larger.

Thereafter, Chomsky’s work is used to present Backus Naur Form notation (Aho et al., 1986). Then, Backus Naur Form is used to specify Context-Free Grammar (Chomsky, 1959), which is commonly used to represent programming-language syntax. Though Context-Free Grammar is not theoretically sufficient for natural language analysis (Joshi et al., 1990), this is often employed for NLP in practice. For example, the Prolog language (Clocksin & Mellish, 2012) was invented for NLP applications to write grammar.

First Order Logic has been one of the most popular representation strategies (Barwise, 1977). It supports syntactic, semantic, and, to a certain degree, pragmatic expressions. The arrangement of the symbols is specified by syntax so that the group of symbols is considered properly formed. The meaning of the properly formed expressions is then specified by semantics. The way contextual information can be leveraged to provide better correlations between different semantics, which is essential for tasks such as word sense disambiguation, is specified by pragmatic.

3.2. *NLP with ML*

Until the 1980s, the majority of NLP systems used complex pre-defined rules. After that, due to the increase of computational power and the shift to ML algorithms, a revolution happened in NLP. Thus, most researches in this field use ML techniques to enhance NLP models. Conventional ML techniques are divided into two categories: supervised learning and unsupervised learning. In un-supervised methods, the task is performed by finding similarities between objects. On the other hand, the supervised models try to learn a function that maps inputs to outputs based on training data. Hence, it can automatically induce linguistic rules from training language data. Hence, the current dominant technique for addressing problems in NLP is supervised learning. Usually, supervised learning approaches are categorized into sequential and non-sequential methods. In the following, a few ML techniques related to NLP are presented. Table 1 shows a summary of these techniques.

Tabel 1. A Few Machine Learning Techniques Used in NLP

Category		Name
Supervised learning	Sequential	Deep Learning
		Conditional Random Fields
		Hidden Markov Model
		Maximum Entropy
	Non-sequential	Support Vector Machines
		Decision Trees
Unsupervised learning		Vector Quantization

- *Deep Learning*: This method is based on artificial neural networks and can extract higher-level features from the raw input to learn multiple levels of abstraction representation (Deng & Yu, 2014). Today, this methodology is very popular for NLP (Mikolov et al., 2013). Using this approach, (Zheng et al., 2013), Chinese word segmentation and Part Of Speech tasks were developed. A deep learning neural network architecture for word level and character level representations, as well as for Part of Speech tagging was proposed by Dos Santos & Zadrozny (2014). Also, a deep neural network system for the information extraction task was proposed in (Qi et al., 2014). Moreover, a deep-learning classifier was evaluated for financial texts sentiment extraction in (Mishev et al., 2019).
- *Conditional Random Field*: Conditional Random Field was introduced by Lafferty et al. (2001). In the NLP domain, Conditional Random Field models are used for many tasks. The Name Entity Recognition (NER) problem in the Chinese language in connection with a Conditional Random Field model was considered in (Yao et al., 2009). Also, different models for language processing based on the Conditional Random Field model for Part of Speech tagging have been proposed (Ammar et al., 2014; Pandian & Geetha, 2009; Patel & Gali, 2008).
- *Hidden Markov Model*: Hidden Markov Model is suitable for text classification. It has achieved great success in text classification, for example, Part of Speech tagging and Name Entity Recognition tasks (Bikel et al., 1999). Hidden Markov Model for the Name Entity Recognition task was proposed in (Morwal & Chopra, 2013). Also, Named Entity Recognition for Hindi, Marathi, and Urdu languages applying Hidden Markov Model was implemented by (Morwal & Jahan, 2013). Furthermore, a rule-based method in combination with Hidden Markov Model was used for Part of Speech tagging (Youzhi, 2009).
- *Maximum Entropy Model*: Maximum Entropy Model is a general-purpose model for making predictions or inferences from incomplete information. Saha et al. (2008) proposed a Name Entity Recognition system based on Maximum Entropy. Additionally, using Maximum Entropy Model, a Part of Speech tagging system was developed by Ekbal et al. (2008).
- *Decision Trees*: Usually, the Decision Tree method is used for the classification task. A method based on Decision Trees for language modeling was proposed by Bahl et al. (1991) to estimate the probability of spoken words.
- *Support Vector Machines*: This approach is used mainly for data classification. In the field of NLP, Support Vector Machines are applied to many tasks, for example, content arrangement, Part of Speech tagging, Name Entity Recognition, and segmentation. A combination of Support Vector Machines and Conditional Random Field has been used for Bengali Named Entity Recognition (Ekbal & Bandyopadhyay, 2009). In Antony et al. (2010) created a Part of Speech tagger for the Malayalam language using Support Vector Machines.
- *Vector Quantization*: The Vector Quantization model organizes data in vectors. This model is suitable for coding applications. For example, this approach was used in speech coding (Makhoul et al., 1985) and audio compression (Spanias et al., 2006).

3.3. NLP with modern ML (Cognitive ML)

Studying the human brain processes for language processing can be beneficial for improving NLP techniques to achieve more human-like behaviors. Additionally, reading a text activates various brain processes that can be recorded by different techniques. Recently, with the evolution of conventional ML techniques, researchers are interested to analyze these cognitive data to develop new ML methods to get a more accurate impression of human language.

For instance, Mishra et al. (2017) used the eye-movement patterns of the reader to derive cognitive features. Also, Hollenstein et al. (2019) showed how adding Electroencephalography (EEG) data yields significant improvements in NLP tasks over the conventional approaches. Moreover, another research used brain imaging techniques to investigate brain activities for reading complex natural text in order to interpret natural language (Toneva & Wehbe, 2019).

Finally, according to the literature, we categorize NLP works into three groups: 1) rule-base works, 2) machine-learning works, and 3) cognitive works. Fig. 2 illustrates these three categories.

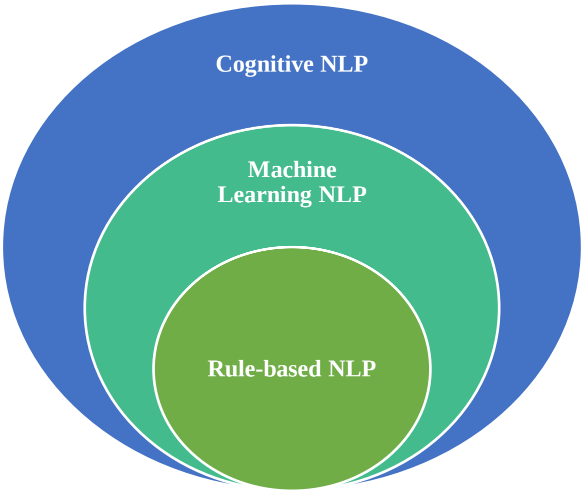

Fig. 2. Three stages of NLP history

4. Discussion and Conclusion

In today's digital era, digital platforms provided new content-sharing services that allow people to communicate easily with each other. Usually, this communication is formed via human language messages. Accordingly, to analyze this big data, we require to computerize natural language contents. For this, the computer science researchers developed several techniques in the NLP framework. Today, NLP is one of the most important modern technologies. In this paper, firstly, the required processing levels for an NLP system to understand a natural language were categorized from low to high levels. Then, we categorized the studied techniques for NLP into three stages: NLP before ML (the rule-based methods), NLP with ML (ML-NLP), and ML with cognitive language processing for NLP (cognitive NLP). The main purpose of this paper was to highlight the latter on as a modern ML technique. The review of these technioques could be beneficial for the interested readers in this topic to create more usable methods.

The essence of NLP is based on linguistics, computer science, and artificial intelligence. Before ML approaches, NLP tasks were commonly carried out using rule-based approaches. In these approaches, rules were constructed manually by linguistic experts or grammarians. However, designing rules required significant human efforts. Despite the fact that most of the presented techniques for NLP are supervised ML, it is expected that unsupervised learning will be popular in the future and this is because human learning is basically unsupervised. In fact, usually, the world structure is normally discovered by observing it, and not abstractly by receiving the names of all objects.

The presence of a large amount of training data affects the performance of ML-based approaches largely. Generally, by increasing or decreasing the size of training data, the performance of ML approaches is enhanced or degraded respectively. Since the free availability of large corpora for most growing languages is a hard issue, a major limitation of NLP today is the fact that most NLP resources and systems are available only for high-resource languages, such as English, French, Spanish, German, and Chinese (in contrast, many languages have low-resources, such as Farsi, Bengali, Indonesian, Punjabi, Cebuano, and Swahili). A future challenge for the language community is how to develop resources and tools for hundreds or thousands of languages, not just a few.

Recently, researchers have been interested to use cognitive language data to take a step towards improving NLP techniques. This approach is necessary to identify cognitive processes of the human brain that are used for language processing tasks. In our opinion, in the future, interactive NLP with cognitive neuroscience will achieve more accurate language processing tools to understand language contents. This direction could be a guideline for developing new ML techniques for major NLP tasks.

REFERENCES

Mankolli, E., & Guliashki, V. (2020, September). Machine learning and natural language processing: Review of models and optimization problems. In *International Conference on ICT Innovations* (pp. 71-86). Springer, Cham.

Aho, A. V., Sethi, R., & Ullman, J. D. (1986). Compilers, principles, techniques. *Addison wesley*, *7*(8), 9.

Al-Harbi, O., Jusoh, S., & Norwawi, N. M. (2017). Lexical disambiguation in natural language questions (nlqs). *arXiv preprint arXiv:1709.09250*.

Ammar, W., Dyer, C., & Smith, N. A. (2014). Conditional random field autoencoders for unsupervised structured prediction. In *Advances in Neural Information Processing Systems* (pp. 3311-3319).

Antony, P. J., Mohan, S. P., & Soman, K. P. (2010, March). SVM based part of speech tagger for Malayalam. In *2010 International Conference on Recent Trends in Information, Telecommunication and Computing* (pp. 339-341). IEEE.

Bahl, L. R., deSouza, P. V., Gopalakrishnan, P. S., Nahamoo, D., & Picheny, M. A. (1991). Context dependent modeling of phones in continuous speech using decision trees. In *Speech and Natural Language: Proceedings of a Workshop Held at Pacific Grove, California, February 19-22, 1991*.

Barwise, J. (1977). An introduction to first-order logic. In *Studies in Logic and the Foundations of Mathematics* (Vol. 90, pp. 5-46): Elsevier.

Bikel, D. M., Schwartz, R., & Weischedel, R. M. (1999). An algorithm that learns what's in a name. *Machine learning*, *34*(1), 211-231.

Bose, R. (2004). Natural Language Processing: Current state and future directions. *International Journal of the Computer, the Internet and Management*, *12*(1), 1-11.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory*, *2*(3), 113-124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and control, 2*(2), 137-167.

Church, K. W., & Rau, L. F. (1995). Commercial applications of natural language processing. *Communications of the ACM, 38*(11), 71-79.

Clocksin, W. F., & Mellish, C. S. (2012). *Programming in Prolog: Using the ISO standard*: Springer Science & Business Media.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing, 7*(3–4), 197-387.

Devasena, C. L., & Hemalatha, M. (2012). *Automatic text categorization and summarization using rule reduction.* Paper presented at the IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012).

Dos Santos, C., & Zadrozny, B. (2014). *Learning character-level representations for part-of-speech tagging.* Paper presented at the International Conference on Machine Learning.

Ekbal, A., & Bandyopadhyay, S. (2009). Named Entity Recognition in Bengali. *Northern European Journal of Language Technology, 1*, 26-58.

Ekbal, A., Haque, R., Das, A., Poka, V., & Bandyopadhyay, S. (2008). *Language independent named entity recognition in indian languages.* Paper presented at the Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages.

Hirschman, L., & Gaizauskas, R. (2001). Natural language question answering: the view from here. *natural language engineering, 7*(4), 275-300.

Hollenstein, N., Barrett, M., Troendle, M., Bigiolli, F., Langer, N., & Zhang, C. (2019). Advancing NLP with cognitive language processing signals. *arXiv preprint arXiv:1904.02682*.

Jackson, P., & Moulinier, I. (2007). *Natural language processing for online applications: Text retrieval, extraction and categorization* (Vol. 5): John Benjamins Publishing.

Joshi, A. K., Shanker, K. V., & Weir, D. (1990). The convergence of mildly context-sensitive grammar formalisms. *Technical Reports (CIS)*, 539.

Jusoh, S., & Al-Fawareh, H. M. (2007). *Natural language interface for online sales systems.* Paper presented at the 2007 International Conference on Intelligent and Advanced Systems.

Kaur, N., Pushe, V., & Kaur, R. (2014). Natural language processing interface for synonym. *International Journal of Computer Science and Mobile Computing, 3*(7), 638-642.

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Liddy, E. D. (1998). Enhanced text retrieval using natural language processing. *Bulletin of the American Society for Information Science and Technology, 24*(4), 14-16.

Liddy, E. D. (2001). Natural language processing.

Makhoul, J., Roucos, S., & Gish, H. (1985). Vector quantization in speech coding. *Proceedings of the IEEE, 73*(11), 1551-1588.

Mani, I., & Maybury, M. T. (2001). Automatic summarization.

Maybury, M. (1999). *Advances in automatic text summarization*: MIT press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality.* Paper presented at the Advances in neural information processing systems.

Mishev, K., Gjorgjevikj, A., Stojanov, R., Mishkovski, I., Vodenska, I., Chitkushev, L., & Trajanov, D. (2019). *Performance evaluation of word and sentence embeddings for finance headlines sentiment analysis.* Paper presented at the International Conference on ICT Innovations.

Mishra, A., Kanojia, D., Nagar, S., Dey, K., & Bhattacharyya, P. (2017). Leveraging cognitive features for sentiment analysis. *arXiv preprint arXiv:1701.05581*.

Morwal, S., & Chopra, D. (2013). NERHMM: A tool for named entity recognition based on hidden Markov model. *International Journal on Natural Language Computing (IJNLC), 2*, 43-49.

Morwal, S., & Jahan, N. (2013). Named entity recognition using Hidden Markov Model (HMM): an experimental result on Hindi, Urdu and Marathi languages. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(4), 671-675.

Pandian, S. L., & Geetha, T. (2009). *CRF models for Tamil part of speech tagging and chunking.* Paper presented at the International Conference on Computer Processing of Oriental Languages.

Patel, C., & Gali, K. (2008). *Part-of-speech tagging for Gujarati using conditional random fields.* Paper presented at the Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages.

Qi, Y., Das, S. G., Collobert, R., & Weston, J. (2014). *Deep learning for character-based information extraction.* Paper presented at the European Conference on Information Retrieval.

Saha, S. K., Sarkar, S., & Mitra, P. (2008). *A hybrid feature set based maximum entropy Hindi named entity recognition.* Paper presented at the Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR), 34*(1), 1-47.

Spanias, A., Painter, T., & Atti, V. (2006). *Audio signal processing and coding*: John Wiley & Sons.

Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). *arXiv preprint arXiv:1905.11833*.

Turing, A. M., & Haugeland, J. (1950). *Computing machinery and intelligence*: MIT Press Cambridge, MA.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks, 5*(1), 7-16.

Yao, L., Sun, C., Li, S., Wang, X., & Wang, X. (2009). *CRF-based active learning for Chinese named entity recognition.* Paper presented at the 2009 IEEE International Conference on Systems, Man and Cybernetics.

Youzhi, Z. (2009). *Research and implementation of part-of-speech tagging based on hidden markov model.* Paper presented at the 2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications (PACIIA).

Zheng, X., Chen, H., & Xu, T. (2013). *Deep learning for Chinese word segmentation and POS tagging.* Paper presented at the Proceedings of the 2013 conference on empirical methods in natural language processing.