

# A Review Paper on Applications of Natural Language Processing-Transformation from Data-Driven to Intelligence-Driven

<sup>1</sup>Aditi Pancholi

B. Tech Scholar, CSE Department  
Indore Institute of Science and Technology  
Indore, India

<sup>2</sup>Ruchir Jain

B. Tech Scholar, CSE Department  
Indore Institute of Science and Technology  
Indore, India

<sup>3</sup>Dhwani Jain

B. Tech Scholar, CSE Department  
Indore Institute of Science and Technology  
Indore, India

<sup>4</sup>Margi Patel

Assistant Professor, CSE Department  
Indore Institute of Science and Technology  
Indore, India

**Abstract**— Natural Language Processing (NLP) is a burgeoning technique used to produce many types of Artificial Intelligence (AI) that we see today, and it will remain a major priority for present and future works for more cognitive applications. In this work, we covered some of the most practical uses of NLP. Our objective is to create a theoretical analysis of various fields where NLP can play a major role and change the whole scenario by its automation techniques. It's a buzzing topic that is attracting everyone to invest in it. These applications are finalized by a refined and thorough study of NLP and its area. This article, beginning with a discussion about trends of NLP and components of Natural Language Processing (NLP), then moving on to the application and emergence of NLP and concerns. We have included Text to Speech generator, Liver cancer prediction, Smart home using IoT, Word Extraction, Sentence Generator, Graph-based methodology, postoperative complications prediction with patients with chronic diseases. NLP has altered our interactions with computers and will continue to do so in the future. As AI technologies modify and enhance communication technology in the years ahead, they will be the underpinning driver for transformation from data-driven to intelligence-driven initiatives.

**Keywords**—Natural Language Processing, Artificial intelligence, Automation, Internet of Things

## I. INTRODUCTION

Natural language processing (NLP) is a research and application field that looks at how computers may be used to interpret and manipulate normal language teaching or speech in order to accomplish meaningful tasks. Common language processing is a hypothetically defined scope of computational techniques that address at least one degree of writing of commonly occurring semantic probes to complete the handling of human-like language for classification of assignments or applications. The purpose of NLP is to carry out clear tasks. Some of NLP's bulk assignments are done with program outlines, talk exams, machine interpretation and so on. NLP helps PCs to connect with human being in their own language and evaluate other language related ventures. For example, NLP makes PCs to understand text, tune to discourse, understand it, measure it, figure out and understand which parts are important.

Current PCs can observe more language-based information than people, with no drawbacks and with a stable, unbiased mind. Keeping in mind the vast measurement of unstructured information that is imported every day from clinical records to online media, computerization will be important to segment text and discourse information as a whole product. The language of humans is unique from that of surprise. We interact in a variety of structures, both orally and in writing. Although there are few accents, each language has an alternate arrangement of punctuation and linguistic structural laws, phrases, and tongues. When we compose, we frequently speak or collect words wrongly or discard utterances. When we speak, we have regional accents, and we mumble, stumble, and learn words in colloquial expressions. Administered and unplanned learning, and particularly intensive learning, are currently widely used to demonstrate human language, moreover a requirement for syntactic and semantic understanding and space mastery that are essentially These AIs do not exist. NLP is important on the basis that it assists with correcting ambiguity in language and for some adds significant numerical design to information, similar to downstream applications, discourse acknowledgments, or text checks.

## II. APPLICATIONS OF NLP

### A. Composing Sentences

Combining linguistic restraints and the model's results can result in meaningful and cohesive phrases. It may be used for a variety of things, including conversation systems, translation software, and picture captioning[1]. It is difficult to inculcate lexical constraints when an auto-regressive model generates sentences left to right through Beam search. BFGAN, a new algorithmic framework, will face this challenging situation. A reverse producer and a forward producer are used in this system to generate coherent sentences using lexical restrictions, and a discriminator is used to combine the backward and forward sentences using signal rewards. Apart than BFGAN, numerous strategies for training ensure that the training procedure more reliable and productive. With time, BFGAN will improve, and other issues will be addressed. Lexically restricted sentence generation[2] is when a lexicon

constraint is a preset set of words that is willing to emerge in the outcome. In the realm of natural language processing, this is the hottest issue right now.

RNN has recently been a prominent player in NLP, with impressive results in tasks such as neural networks [3], tech blog creation [4], textual summarization [5], table-to-text creation [6], and effective text production [7]. Beam Search (BS) is conducted and left to right sentences are generated using auto-regressive models[8]. It is a difficult process to generate sentences that contain lexical constraints. The fluency of the sentence will be damaged if we replace the arbitrary word in the output. There is no assurance that the desired term will appear in the outputs if extra information about the term is provided [9], [10]. Reverse and forward linguistic models, which function together, are used to construct lexically limited phrases in (B/F-LM). The reverse linguistic model creates the very first part of the sentence, with the input being a lexical constraint. The forward linguistic model then creates the entire sentence by taking the first half-sentence. These two models are trained using maximum likelihood estimation (MLE) objectives[1]. When the reverse language model created the first half and the forward language model created the back half, the outputs were inconsistent and unintelligible. Because both language models were trained separately, this issue arose, they should be trained together and they should have the access to one another's output. To solve the challenges of lexically confined languages, a novel approach called reverse-Forward Generative Adversarial Network (BFGAN) was developed. Reverse producer, forward producer, and classifier are the three components. The two generators work together and fool the discriminator. The dynamic attention mechanism is introduced to the forward generator for coherence. A recursion takes place from beginning to the end and the scope of the attention function increases. During inference, the forward producer can attend to the first split formed by the reverse producer[1]. BFGAN addresses the issues and improves the working of lexically constrained language and algorithms are introduced to ease the working of the model which also provided stability to it. Future works will include working on multiple lexical constraints.

#### *B. Prediction of postoperative complications*

Hospital readmissions occur due to the failure of hospital laws and orders related to the discharge, recurrence of infections, poor care facilities after the discharge from the hospital. This also may occur due to the irresponsibility or carelessness of the hospital staff. The federal government has imposed several rules and regulations. There are several needs of the patient after discharge, extensive care coordination should be there for COPD patients, and post-discharge needs should be taken care of, but readmissions after 30 days would be by the cause of chronic pulmonary diseases. United States government imposes a penalty on the hospitals having excessive readmissions. Predictions are dependent on the clinical notes. Based on medical history, allergies, demographics, lab results reports are being made and analyzed. It is a high maintenance problem that needs to be handled cautiously and responsibly. Hospital readmissions can be avoided by taking care of those patients who are prone to the risk of readmission. ACO's and MSO's are targeting hospitals to improve their conditions about readmissions. Raw data is mainly unstructured which

has patient's data consisting of physician's notes, discharge notes, laboratory reports, and all of this are also included in predictive analysis. But there is a problem, that is every doctor has their style of writing such as acronym and different terminologies or abbreviations, which arises ambiguity and complexity to resolve, So a system is required to convert unstructured data to structured data. On the other side, NLP is accomplishing major tasks and is becoming successful. Several libraries, frameworks, and toolkits are available to save valuable minutes. Rather than building everything from scratch, using these NLP libraries so much of time will be saved because it contains many classes, subroutines, snippets, functions, etc. Machine learning approaches are used to train these models. Frameworks also come into the picture and the similarity between frameworks and libraries is that both help in the reusability of code. OpenNLP and java-based NLP library is used in this predictive model. OpenNLP's purpose is to get a library for very well NLP applications such as cryptocurrencies, reducing vulnerability, portion tagging, named entity recognition[11]. Frameworks are useful for NLP processing since the pipeline pattern [12] is a well-known design paradigm in the field. NLP tasks are frequently organised from low-level to high-level, with each query perhaps reliant on the preceding. Users can construct a pipeline of actions tailored to the system's aim. Many versions of NLP tools that use the pipeline model structure have stood the test of time. These routines are known as annotators. Medical notes, laboratory reports, course treatments, allergies, medicines, genetic medical conditions, and other important documented information are taken into observation but still, it does not give us a clear picture of the patient's health condition, they are just for billing purpose, discharge formalities and for various other hospital-related laws. These unstructured pieces of information are to be converted into structured information through the predictive model and there is a reason why these documents are not structured, as they would require special extra staff for this purpose mainly known as medical coders. Unstructured data from the Electronic Medical Record is being used to predict 30-day readmission in patients with chronic diseases[14]. Despite the fact that research utilising NLP to predict patient outcomes has begun to surface in the previous years, no study has evaluated feature selection strategies for patient outcomes, and only a few publications are being done in predicting patient outcomes [13]. Many clinical notes from patients having chronic diseases from previous years have been used in the software. The patients which are seriously diagnosed with chronic diseases are included in the data. We identify the features and track the patient's readmission status within 30 days following discharge from the very same medical facility. Extraction of features takes place from the dataset. As the quality of the feature increases, the quality of the model also increases. As a result, an NLP approach is made to predict hospital readmissions. By using naïve Bayes classifier with chi-squared the feature became more effective. Using the bag-of-words approach and UML, we can easily predict hospital readmissions. The point of advantage is that the information related to the patients like medicals notes, history, reports, etc. is already collected by the medical centers. RDBMS is used and easy incorporation into the existing electronic medical

system. Clinical notes will become more relevant with increase in electronic medical systems, and NLP approaches will need to be addressed when developing prediction models. The analysis indicated the relevance of feature selection and model building time for practical system deployment. As more records become accessible, future attempts will be expanded to include other chronic conditions[11].

### C. Word Extraction

As the area of NLP is growing faster, it is expanding its horizon. Word Extraction is also possible due to NLP. We can generate associated words from the given words and collect them. This replaced the manual collection of words, This application is an autonomous extraction of connected words approach. It has two stages association word extraction and machine association network construction. The reading comprehension algorithm is used mostly for the extraction of words automatically. Association is being connected to some person or a thing and cue response is a signal message in response to some activity and behavior. This is very time-consuming and an expensive way of word association tasks. Various experiments are performed under environmental and experimental conditions which are very different from real life. Studies are still going on in the NLP sector because the cue response word association is not that effective and accurate. By (RC) Reading Comprehension algorithm of NLP associative words are found. There is a machine association network in which words are machine-generated and on the other hand, there is a human association network in which words are manually collected. Based on the attention model, words are extracted. The attention mechanism is used for figuring out the relationship between cue-response words. With the help of plain texts, associated words are extracted and a network is a build-up. NLP has the capability of identifying, understanding, and analyzing the relationship between words linguistically and semantically. Mapping is done from cue-word to response word by reading comprehension algorithm(RC). The human association pattern is stable to some extent. Coffee- Brazil, Coffee-Rocket, So we expect the machine association network to be accurate to some extent without varying the attention entity[15]. Regardless of the attention algorithm utilised, the attention mechanism is known for establishing reasonably persistent cognitive patterns. This leads us to the conclusion that this neural network-based framework gathers connected terms from text format, as well as a Reading Comprehension algorithm and an attention mechanism are used to complete the task.

### D. NLP pipeline for liver cancer prediction

Despite the growth in the NLP field, EMR processing works remained scarce and challenging due to the scarce information and datasets and semantical features, especially for radiology reports. This research, the NLP pipeline is used to extract features from laboratory reports. Cancer risk prediction is also predicted. For Liver cancer speculation, the random forest has best performance with good accuracy. NLP pipeline can also be used with clinical tests and other predictive tasks. EMR's electronic medical data are precious assets. This application focuses on radiology reports

and cancer predictions. The NLP pipeline for feature extraction could be incorporated in other types of clinical context and disease speculation engagements. For improved quality of care and coordination, electronic medical reports are significant in the field of research. Machine learning-based techniques play a significant role in data analysis in today's digital era, and are beneficial in fields such as medical decision, disease treatment, and administration [16], [17]. The major means of communication between radiographers who scan photos and medics who write the final control study is through medical images and lab results . NLP utilizes mathematical and linguistic approaches to extract information from unstructured information, which is then turned into large datasets. The advantages of NLP-based feature extraction are numerous. As a result, NLP-based feature extraction has been successfully applied to diagnostic monitoring, cohort creation, performance evaluation, and clinical assistance in radiology [18]-[21]. Free text is found more relevant and natural than in a clinical scenario, here NLP is used to facilitate the extraction of features, entities, relations from the clinical texts. A sequence of labeling tasks is called named entity recognition (NER). It's a complex process of extraction of information from EMR's. Deep neural networks were recently introduced to NER in order to boost effectiveness. Tasks such as drug-related studies, disease studies, etc, in these tasks feature extraction, are implemented. Lasso is introduced for feature selection. Further, these test reports use feature selection by lasso and binomial distribution of logistic regression is used. Liver cancer is a implication for both the sufferer and the country's regime. Many patients are diagnosed with terminal liver cancer when the situation and the health conditions go out of hand. Therefore, early diagnosis of liver cancer should be there. For EMR processing of radiology reports, limited corpus and datasets are present, therefore we have to manually make the corpus for our study. A lexicon is constructed which is collected by the radiological experts and annotation is done to the same, based on the Linguistic constraints and clinical experience of the experts these are processed. Clinically related words are considered by this lexicon. Because we only used a tiny percentage of radiological reports to create the lexicon, the pipeline may be employed as a reference in similar Electronic medical applications[22]. Radiological features can be extracted by the pipeline and the advice can be given to the physicians by the diagnosis model. The investigation is carried out about how neural network-based scheme is helping radiologists to make decisions about the diagnosis especially radiologists with limited experience. The ability of the diagnosis scheme has improved a lot. The pipeline is also showing high-performance accuracy in liver cancer diagnosis. NLP pipeline is described in this study for liver cancer diagnosis. To boost the effectiveness of the NER BiLSTM-CRF a deep learning lexicon model is incorporated in this study, the model gives accurate results with both NER and liver cancer prediction. The suggested NLP process might be used to the creation of lexicons for additional illnesses and types of medical texts[22].



### E. Interactive text2speech

Language is one of the most efficient ways of communication but it can create uncertainty and ambiguity and can be a source of misinterpretations. A person may make ambiguous statements that can be construed in a variety of ways[23]. So without linguistic interpretation and object recognition, Interactive Text to Pickup can accomplish the given task despite the vagueness in human commands and restrictions. Linguistic interpretation is the process of analysing a string of symbols, whether in natural language, code generator, or database systems, using the rules of a formal grammar, syntax analysis, or syntactical analysis, and Object Recognition is a computer vision method that recognizes and determine objects in images or videos, can be utilized to count the objects and trace their precise location. We eliminate the necessity for the pre-processing stage, as well as linguistic interpretation and object identification, by creating the Text to pickup network in an end-to-end way. When an input instruction is difficult or several items are arranged in different ways, the prepared Text to pickup network can manage the task flexibly[24]. Interaction is the solution to remove ambiguity. It is a system that consists of the text 2 pickup network and a question generator. Furthermore, the Text to pickup network produces a heatmap of position. Position heatmap is a data visualization method that shows the magnitude of a phenomenon as color in two dimensions. It's a two-dimensional distribution with values indicating assurance in the target object's position, as well as a heatmap of uncertainties. The initial language instructions by the user will be ambiguous and the question generator network will determine which question to ask the user. It is a very challenging task to determine the question to be asked because based on this further the process is continued. The question needs to be relevant and to the point and well-defined, it should not overlap the previously given information.[3] Once we get the response to the additional question requested, the answer is then annexed to the primary command and it goes back to the Text to pickup network, in short, it is a visual question answering (VQA). Note that producing a question to reduce the uncertainty of image-related linguistic set of instructions and finding a response to a given image-related inquiry are two separate jobs.[23] Using the Baxter robot this network illustrates the functioning, (IT2P) for its training and testing, a dataset is collected based on the images taken by the Baxter robot. On the arm of the robot, the camera is fitted. Each image contains 3 to 6 blocks of fine colors. The purpose of fixing the camera was to make the robot acquainted with the real-world environment.

Commands will be the incorporation of representations and guidance, firstly associated to the position (leftmost, right, topmost, left), associated to color (red, yellow, blue, white), associated with the relative position (between two purple blocks).

The accuracy before and after the interaction is compared. Interaction played a significant role here by eliminating the ambiguity from the commands. Interactive Text to pickup network and a single Text to pickup (the one with no interaction) are compared. Here in the interactive Text to pickup network, we have a simulator that informs the color

and position of the block. Its accuracy is 86.68% and the single Text to pickup network has an accuracy of 45.66%.[24]

For instance, the network asks "This one", if Yes then the simulator says Yes, If no the experiment is considered unsuccessful. For making this experiment more rigorous and realistic if the command says "pick up the red one.", the network says "red one?". we propose the Interactive Text to Pickup (IT2P) network for retrieving the sought object in response to a human language command. When an unclear language command is given, the IT2P network engages with an user interface to resolve the uncertainty. The suggested network can correctly anticipate the position of the intended item and the uncertainty associated with the expected target location by learning the supplied language command. The suggested IT2P network is proven to be capable of efficiently interacting with persons by posing a question that is relevant to the circumstance. The suggested network has been tested on a Baxter robot, as well as collaboration between a real robot and a person. We think that the suggested system can effectively communicate with humans by asking questions based on estimating uncertainty, allowing for more straightforward human-robot collaboration.

### F. Graph based method for NLP and NLU

This paper provides the functional component, performance, and maturity of graph-based methods for natural language processing and natural language understanding. Extracted proficiency from the methods includes summarization, text entailment, redundancy reduction, similarity measure, labelling. From all these methods approximate accuracy, coverage, scalability and performance are derived. This survey and analysis, with tables and graphs, offers a unique extraction of functional components and levels of maturity from this collection of graph-based methodologies.

The vastness of data combined with the necessity for quick access to specific but comprehensive information has driven natural language processing (NLP) and natural language understanding (NLU) research to supply the subsequent capabilities: event resolution (ER), grammar annotation (GrA), information mining (IM), knowledgebase (K), labelling (Lab), novelty recognition (ND), question/answer (QA), redundancy reduction (Red), semantic relatedness (SR), similarity measure (SM), summarization (Sum), textual entailment (TE), word sense disambiguation (WSD), and word sense induction (WSI)[25]. It is found that the graph-based methods have less complexity than the vector method which offers a more compressed and efficient concept of representation of text. The goal is to supply the reader with almost every detailed information together with tables and charts to capture the present state of the art in graph-based methods for NLP and NLU, including their component functions, performance and maturity.

Following describes each of those areas as NLP and NLU capabilities:

1.Summarization captures the most meaning from the documents and also down it to a specific level of detail.

2.Textual Entailment, at the syntactic level, replaces all subsets of the text, with the encompassing text. At the semantic level, it can be understood, that the truth of a

sentence is acquired by the reality of another sentence without losing its meaning. Through the required result's within the variety of a shorter summary and also with the initial meaning.

3. By measuring the similarity of text or their concepts, the resulting similarity measure can be used to merge the clusters of comparable concepts into a single concept.

4. Semantic relatedness is the relation between concepts. Concepts, relations, and attributes are used to represent text at a semantic level and higher-level abstractions of the meanings represented by a body of text. A potential goal for such higher level (or semantic) representation is natural language understanding (NLU)[26].

5. Labelling is also used to represent the parts of speech and senses of words within a text. Labelling are often generated manually, supervised, semi-supervised, and unsupervised methods. Supervised learning methods are trained using a manually labelled collection of texts. Semi-supervised methods are trained using some amount of texts to derive labels for big unlabeled texts. Unsupervised methods are trained with a much larger collection of text. These learning methods are used to determine the sense of words in an unlabeled text.

6. Novelty recognition uses some of the above-mentioned capabilities to detect events of interest within a text.

#### G. Smart Home implementation

In the light of this modern era, the intention of the modern Internet of Things (IoT) is to provide services based on the systems which are completely trained by keeping the user's convenience and accuracy in mind. Various researches are also underway to create smart environments such as a smart house, grids, and industrial IoT environments. Smart homes which are based on IoT technologies include IoT features that are beyond the usual/ordinary network technologies. Miscellaneous technologies of IoT are needed to build an expert system in the surrounding area which requires continuous management, such as IoT-based smart home systems, or to support device and monitoring services and make environments that are suitable to the user through communications between devices to enhance domestic lifestyle services [27]. Real-time sensor data is undertaken as an essential component of an IoT environment to provide services that are customized for the user. The number of sensors is needed to gather the data in an IoT environment and processing method for the collected data through a sensor network. When we talk about the conventional home system in that the user changes the environment manually by controlling devices that are connected to the same network via a mobile device which also leads to data waste and power loss when tasks are executed without the specific characteristics of a user. However present-day IoT based home systems use a method that operates devices intelligently according to the threshold values specified to control the environment [28]. During this study, multiple MJoin operators are accustomed to efficiently process sensor data (stream data) in an IoT environment. Also, a global shared query technique used for the optimized query, the Support Vector Machine classification algorithm was used to distinguish and minimize the data to enable structured storage

management. IoT environments are composed of three technologies: sensing technology, which measures changes within the environment; interface technology, which performs or links bound options through folks, things, and services; and network infrastructure technology, which creates networks between sensors and services [29]-[30]. These applied sciences provide a list of environments, comprising remote control, which performs according to the user's needs and automatic controls, which recognizes people and provides custom services. Join queries are required to process comprehensive data that are acquired from not only one sensor, but multiple sensors, as within the case of an IoT environment. Join operators include operators based on hash tables, windows, and both hash tables and windows [31]. The conventional home system has drawbacks of executing single which lead to various losses, so to perform multiple tasks IoT environment is set up with more process data and also set priorities among tasks such that high-priority tasks are processed. The SVM classification algorithm is used to reduced and classify the data, then this data is saved to TinyDB. Real-time and non-real-time are segregated in the server. Real-time tasks are provided to the user and then saved in the MainDB. Non-real time data are saved within the MainDB and provided if requested by the administrator. An Arduino (Uno) module is used as the processor board, and five sensors are used to acquire streams (temperature, humidity, gas, vibration, and recognition) for sensor data processing. All these data originate from the same environment. Therefore, they are combined into one packet and transmitted. Each packet generates additional traffic and consumes energy. Therefore, a single packet is used to process a query [32]. To control the smooth sensor flow, data generated by the sensors and user task commands are required by the system to check the frequency and time of resultant data. Sensors with low usage rates are changed to standby or drop state, based on the analysis of usage rate in accordance with month, week and day[33]. Users use applications to set the sensor threshold value range, and a device can be controlled by checking the sensor data. The range values set in the database are loaded and referred to when performing tasks. The sensors consist of temperature, humidity, gas, recognition, and vibration sensors in the home. The temperature sensors were designed to be combined with the gas sensors such that they can handle dangerous situations. Additionally, tasks performed with several sensors are digitized and stored within the database. Supported the sensor data, the system provides a convenient IoT environment by performing tasks within which the events that are associated with the tasks are suitable for things or the environment. Based on the sensor data range set by the user and the database data, the user analyzes the data in the application to control the sensors. Sensors with low usage can be placed in a standby state, and a sensor's state can be changed to be operational again [33]. In addition, the priorities among tasks are set such that the number of simultaneous tasks can be reduced by suspending existing tasks when a higher-priority task occurs, to reduce unnecessary and wasteful electricity usage while providing intelligent services to the user.

## CONCLUSION

Natural Language Processing is one of the hottest topics in computer science right now. Companies are investing a lot of money in this field's science. All is attempting to grasp Natural Language Processing and its applications in order to pursue a career in this field. Any company needs to incorporate it into their operations in some way. In this paper we read few applications of NLP but there are many more in the list.

## REFERENCES

- [1] BFGAN: Backward and Forward Generative Adversarial Networks for Lexically Constrained Sentence Generation Dayiheng Liu, Jie Fu, Qian Qu, and Jiancheng Lv
- [2] C. Hokamp and Q. Liu, "Lexically constrained decoding for sequence generation using grid beam search," in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 1535–1546.
- [3] Y. Wu et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," 2016, arXiv preprint arXiv:1609.08144.
- [4] L. Dong, S. Huang, F. Wei, M. Lapata, M. Zhou, and K. Xu, "Learning to generate product reviews from attributes," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 623–632.
- [5] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 1073–1083.
- [6] T. Liu, K. Wang, L. Sha, B. Chang, and Z. Sui, "Table-to-text generation by structure-aware seq2seq learning," in Proc. AAAI Conf. Artif. Intell., 2018, pp. 4881–4888.
- [7] S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency, and S. Scherer, "Affect-LM: A neural language model for customizable affective text generation," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 634–642.
- [8] F. J. Och and H. Ney, "The alignment template approach to statistical machine translation," *Comput. Linguistics*, vol. 30, pp. 417–449, 2004.
- [9] T.-H. Wen et al., "Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking," in Proc. Annu. Meeting Special Interest Group Discourse Dialogue, 2015, pp. 275–284.
- [10] J. Yin, X. Jiang, Z. Lu, L. Shang, H. Li, and X. Li, "Neural generative question answering," in Proc. Int. Joint Conf. Artif. Intell., 2016, pp. 2972–2978.
- [11] A Natural Language Processing Framework for Assessing Hospital Readmissions for Patients With COPD Ankur Agarwal, Christopher Baechle, Ravi Behara, and Xingquan Zhu
- [12] D. Ferrucci and A. Lally, "UIMA: An architectural approach to unstructured information processing in the corporate research environment," *Nat. Lang. Eng.*, vol. 10, no. 3/4, pp. 327–348, 2004.
- [13] A. Agarwal, R. S. Behara, S. Mulpura, and V. Tyagi, "Domain independent natural language processing – A case study for hospital readmission with COPD," in Proc. 2014 Int. Conf. IEEE Bioinf. Bioeng., 2014, pp. 399–404.
- [14] J. H. Wasfy et al., "Enhancing the prediction of 30-Day readmission after percutaneous coronary intervention using data extracted by querying of the electronic health record," *Circulation Cardiovascular Quality Outcomes*, vol. 8, no. 5, pp. 477–485, 2015.
- [15] A Natural Language Process-Based Framework for Automatic Association Word Extraction Received November 21, 2019, accepted December 12, 2019, date of publication December 30, 2019, date of current version January 6, 2020.
- [16] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, no. 6, pp. 395–405, May 2012, doi: 10.1038/nrg3208.
- [17] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, Jan. 2015, doi: 10.1016/j.csbj.2014.11.005.
- [18] D. J. Goff and T. W. Loehfelm, "Automated radiology report summarization using an open-source natural language processing pipeline," *J. Digit. Imag.*, vol. 31, no. 2, pp. 185–192, Apr. 2018, doi: 10.1007/s10278-017-0030-2.
- [19] H. T. Huhdanpaa, W. K. Tan, S. D. Rundell, P. Suri, F. H. Chokshi, B. A. Comstock, P. J. Heagerty, K. T. James, A. L. Avins, S. S. Nedeljkovic, D. R. Nerenz, D. F. Kallmes, P. H. Luetmer, K. J. Sherman, N. L. Organ, B. Grifith, C. P. Langlotz, D. Carrell, S. Hassanpour, and J. G. Jarvik, "Using natural language processing of free-text radiology reports to identify type 1 modic endplate changes," *J. Digit. Imag.*, vol. 31, no. 1, pp. 84–90, Feb. 2018, doi: 10.1007/s10278-017-0013-3.
- [20] S. K. Kang, K. Garry, R. Chung, W. H. Moore, E. Iturrate, J. L. Swartz, D. C. Kim, L. I. Horwitz, and S. Blecker, "Natural language processing for identification of incidental pulmonary nodules in radiology reports," *J. Amer. College Radiol.*, vol. 16, no. 11, pp. 1587–1594, Nov. 2019, doi: 10.1016/j.jacr.2019.04.026.
- [21] K. L. Kehl, H. Elmarakeby, M. Nishino, E. M. Van Allen, E. M. Lepisto, M. J. Hassett, B. E. Johnson, and D. Schrag, "Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports," *JAMA Oncol.*, vol. 5, no. 10, p. 1421, Oct. 2019, doi: 10.1001/jamaoncol.2019.1800.
- [22] A Natural Language Processing Pipeline of Chinese Free-Text Radiology Reports for Liver Cancer Diagnosis HONGLEI LIU 1,2, YAN XU3, ZHIQIANG ZHANG1,2, NI WANG 1,2, YANQUN HUANG1,2, YANJUN HU3, ZHENGHAN YANG3, RUI JIANG4, AND HUI CHEN1,2
- [23] S. L'obner, *Understanding Semantics*. Evanston, IL, USA: Routledge, 2013.
- [24] Interactive Text to pickup Networks for Natural Language-Based Human-Robot Collaboration by Hyemin Ahn, Sungjoon Choi, Nuri Kim, Geonho Cha, and Songhwa Oh.
- [25] Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis Michael T. Mills and Nikolaos G. Bourbakis, Fellow, IEEE
- [26] G. Ambwani and A. R. Davis, "Contextually-mediated semantic similarity graphs for topic segmentation," in Proc. Workshop Graph-based Methods Natural Language Process., Jul. 2010, pp. 60–68.
- [27] M.-Z. Song, "A study on business types of IoT-based smarhome: Based on the theory of platform typology," *J. Inst. Internet, Broadcast. Commun.*, vol. 16, no. 2, pp. 27–40, 2016, doi: 10.7236/JIIBC.2016.16.2.27.
- [28] M. Alaa, A. A. Zaidan, B. B. Zaidan, M. Talal, and M. L. M. Kiah, "A review of smart home applications based on Internet of Things," *J. Netw. Comput. Appl.*, vol. 97, pp. 48–65, Nov. 2017, doi: 10.1016/j.jnca.2017.08.017
- [29] D.-W. Song, K.-S. Kim, and S.-K. Lee, "An relational analysis between humidity, temperature and fire occurrence using public data," *Fire Sci. Eng.*, vol. 28, no. 2, pp. 82–90, Apr. 2014, doi: 10.7731/KIFSE.2014.28.2.082
- [30] H. Yang, "The novel modern Internet of Things system structure optimization methodology based on information theory and communication signal transmission model," *Int. J. Future Gener. Commun. Netw.*, vol. 9, no. 9, pp. 119–132, Sep. 2016, doi: 10.14257/ijfgen.2016.9.9.11.
- [31] R. M. Duarte, A. R. Du Bois, M. L. Pilla, G. G. H. Cavalheiro, and R. H. S. Reiser, "Comparing the performance of concurrent hash tables implemented in haskell," *Sci. Comput. Program.*, vol. 173, pp. 56–70, Mar. 2019, doi: 10.1016/j.scico.2018.06.004.
- [32] J. Yu, J. Li, Z. Yu, and Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 15, 2019, doi: 10.1109/TCSVT.2019.2947482.
- [33] Design of Smart Home Implementation Within IoT Natural Language Interface TAE-YEUN KIM 1, SANG-HYUN BAE 2, AND YOUNG-EUN AN