

The Progress That Natural Language Processing Has Made Towards Human-level AI

Dantong Zhang^{1,*}, Jingwenqi Wang² and Meixuan Sun³

¹*Department of Humanities and Creative Writing, Hong Kong Baptist University, Hong Kong 999077, China*

²*Faculty of Business Administration, Nebrija University, Madrid 28015, Spain*

³*Faculty of Journalism and Communication, Communication University of China, Beijing 100024, China*

*Corresponding author: 850945300@qq.com

Keywords: future of artificial intelligence, natural language processing (NLP).

Abstract: First of all, in this paper we will briefly introduce the development of Natural Language Processing (NLP) over time and then discuss how this technology benefit our humans, like what initial problems which beat human are being solved by this technology. We also summarise how this technology is being used upon various domains. Last but not least, we predict the future development of NLP by putting forward detailed understanding of the past and present in the context of artificial intelligence.

1. Introduction

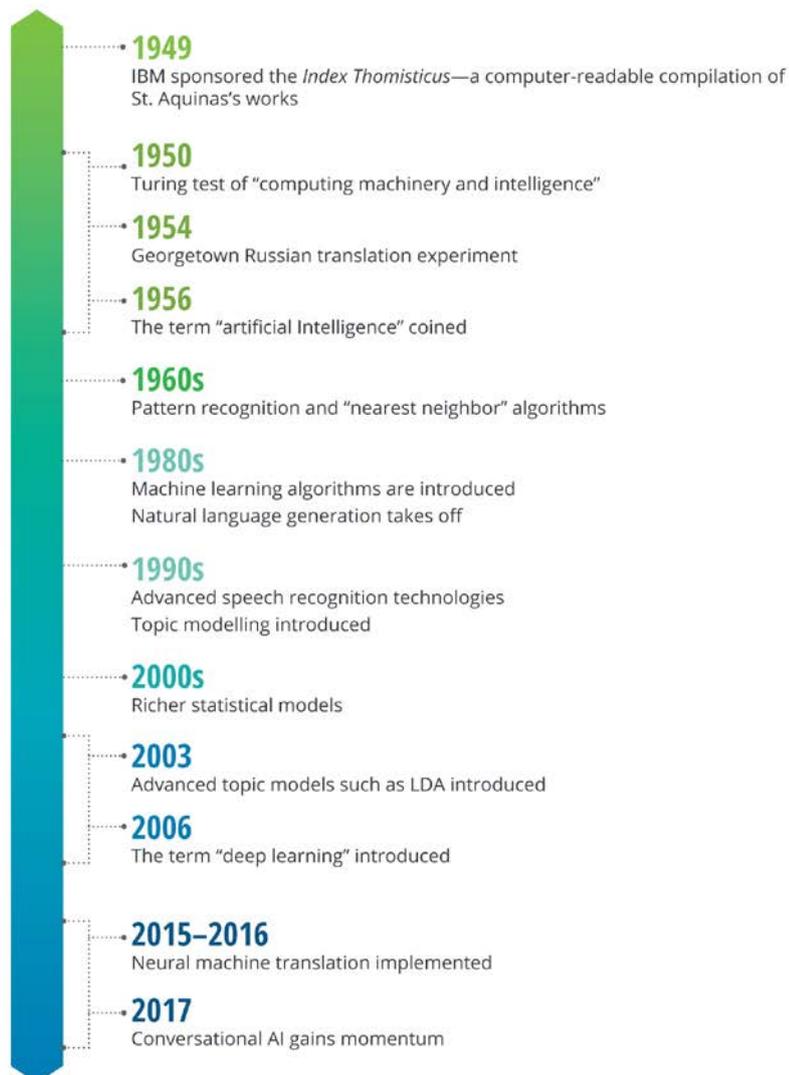
Natural Language Processing (NLP) usually rely on machine learning to derive meaning from human language in an AI discipline, and it is considered to be difficult in computer science. This is because understand human language requires both understanding of how the words and concepts are connected with one another in order to deliver the intended message, it may focus on some specific contexts to convey different meanings (Garbade, 2018). The ultimate goal of NLP is to make sense of human languages in a way which is valuable. In this paper, we will provide an overview of the development of NLP, to see how it has transformed, refined and strived for human-like performance as well as looking into the how it can be used and achieved. With the techniques used in this technology, many issues in our daily lives has been addressed in a good manner, from common applications to the context of social culture, healthcare and business.

2. Brief History of NLP

Natural Language Processing can be traced back to the late 1940s. The first computer-based application related to NLP is Machine Translation (MT). In 1946, Weaver and Booth started one of the earliest MT projects, it is based on expertise in breaking enemy codes in World War II, it is inspired by Weaver's idea of memorandum of 1949 which raised MT to public attention. He also suggested to use the idea from information theory and cryptography for language translation. Soon,

the research about the subject began in the United States within different institutions (Liddy, 2011).

The evolution of NLP and underlying algorithms



Sources: Thomas N. Winter, "Roberto Busa, S.J., and the inventor of the machine-generated concordance," University of Nebraska-Lincoln, January 1999; Bhargav Shah, "The power of natural language processing: Today's boom in artificial intelligence," *Medium*, July 13, 2017; Chris Smith et al., *The history of artificial intelligence*, University of Washington, December 2006; Eric Eaton, "Introduction to machine learning," presentation, University of Pennsylvania; Kendall Fortney, "Pre-processing in natural language machine learning," *Towards Data Science*, November 28, 2017; Clark Boyd, "The past, present, and future of speech recognition technology," *Medium*, January 10, 2018; Regina Barzilay, "Natural language processing," lecture, MIT, September 2012; Build with Google Cloud, "A history of machine learning," accessed December 19, 2018; Thomas Hofmann, "Probabilistic latent semantic indexing," proceedings of the twenty-second Annual International SIGIR Conference on Research and Development in Information Retrieval, accessed December 19, 2018; Robert Dale, Barbara Di Eugenio, and Donia Scott, "Introduction to the special issue on natural language generation," *Computational Linguistics* 24, no. 3 (September 1998); *Medium*, "History and frontier of the neural machine translation," August 17, 2017; Ram Menon, "The rise of the conversational AI," *Forbes*, December 4, 2017.

Figure. 1 Deloitte Insights

The early work in MT took a simplistic approach that only differences between languages resides in their vocabularies and word orders (Liddy, 2011). Because of the development of the syntactic theory of language and parsing algorithms, NLP received widespread recognition the 1950s. There was an over-enthusiasm among groups of people who believe that the system of automatic high-quality translation can produce results just as indistinguishable as human translators (Liddy,

2011). But apparently, this idea is unrealistic given the limited linguistic knowledge. This also led to the report of 1966 from the Automatic Language Processing Advisory Committee (ALPAC) of the National Academy of Science. The report indicated that the MT could not be achieved in a short-term.

In spite of the fact that there was a decrease in NLP work after the ALPAC report, some developments turned out to be significant. In the late 1960s and early 1970s, theoretical work focused on the issue of how to develop computationally solutions that the then-existing theories of grammar could not produce. In late 1970s, the attention shifted to communicative plans, discourse phenomena and the semantic problems. Some great contributors including Polanyi and Scha, Hobbs and Rosenschein and Reichman. By the end of 1980s, NLP researchers used symbolic approaches which had been used to address different problems. In the 1990s, a historian called Sharon Block used *topic modeling* to conduct quantitative analysis of the Pennsylvania Gazette, which became one of the most popular newspapers in the US during 18th century (Eggers, Mailk, & Matt, 2019).

In the last ten years of the millennium, NLP was growing fast. As Eggers, Mailk and Matt (2019) pointed out: “The biggest advances in this field seem to be guided by deep-learning methods based on neural networks. The methods are designed to mimic the function of the neurons in a human brain to ensure better performance. Unlike applications based on algorithms methods, deep-learning NLP can automatically learn from examples.”

3. NLP can help us to address different issues

Natural Language technologies is the driving force behind various applications. The following applications has helped us to solve problems in real-life and improve our living standards:

- The use of machine translation, text processing and language generation (e.g. Google Translate).
- Speech to text and text to speech.
- Word Processors like Grammarly employs NLP to check grammatical accuracy of the texts.
- In call centers, the Interactive Voice Response (IVR) application is used to respond to user’s request.
- Personal assistant applications (e.g. Siri, Alexa and Cortana).

Key NLP capabilities

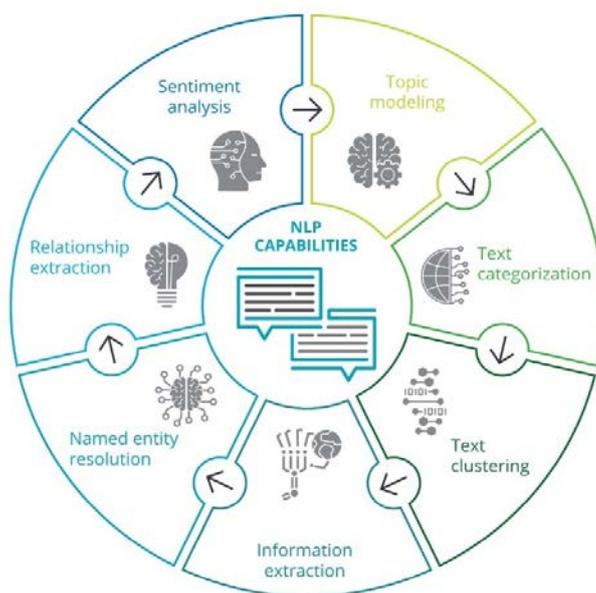


Figure. 2 Deloitte Analysis

What is more, NLP provide the tools that the government may need for improve public service, identify potential risks and even solve crimes. Some researchers from Claremont Graduate University found that this technology can comb through news articles to identify vital elements like people, time and locations with high precision (Eggers, Mailk, & Gracie, 2019).

There are several ways that NLP can help to address various issues for government agencies. It can analyze the feedbacks and comments from the internet users, especially the unstructured contents, which is more efficient than our humans. It can help government agencies to understand citizens' concerns in order to serving the public in a better way and provide answers to their citizens' questions. Singapore created "Ask Jamie" by using NLP, it is trained to respond to the users on agency websites, as a virtual assistant, Ask Jamie can also raise questions to narrow down answers which are relevant to the query posted online (Eggers, Mailk, & Gracie, 2019). The technology can also help to enhance policy analysis and facilitate predictions to design measures. In Peru, a place where the discussions on public service and infrastructure has been going on for years, by using the help from *topic modeling*, policy makers can establish a negative correlation between policy volatility and its development in the long run (Eggers, Mailk, & Gracie, 2019).

3.1 The use of NLP in Business, Media and Healthcare Areas

NLP can also be used in business, media and healthcare areas. The chat robot uses artificial intelligence open source software packages and artificial intelligence natural language processing cloud service technology, which can recognize real human-like interaction and emotional embedding of the chat robot. At the same time, chatbots need natural learning and processing to quickly identify the semantics and context of the text. Without the technology of natural learning processing parser, the accuracy of computer recognition of humanization and understanding of multi-scene topics is low.

NLP is applied to the Internet to make corresponding user-related needs. The technical tools needed are a combination of natural language processing and statistics. The basic theory is to assign values (positive, negative or neutral) to texts and then process and retrieve according to the values assigned to the text to create a human perception that can link the previous and the next (emotion, thinking, reaction).

NLP uses the keyword matching technology of NLP in the text for advertising in marketing. Keyword matching is a basic but highly practical NLP technology. The intelligent assistant used in diversified scenarios is a technical application that is based on NLP to achieve processing tasks through CUI (CUI-Conversational User Interface) on a computing device and presented to users in a conversational form for human-like interaction. Such as chatbot, service bot and P2P e-commerce.

NLP was successfully applied to the processing technology of the decision support system used in patient cohort analysis. It is used in detail in the establishment of the knowledge base, language processing, clauses, word segmentation, semantic analysis, text summarization and the final clinical decision support system.

According to Gennaro Cuofano, the creator of FourWeekMBA, NLP technology can be used in social media in a powerful way. The Chinese creative social media platform TikTok have its research lab, what the company ByteDance named "AI Lab". It claims expertise in many areas. For example, in NLP filed, the company creates applications like:

Margin-bottom-22">

Table 1 Applications ByteDance creates

Byte Translator	Machine translation service for ByteDance Products
Xiaomingbot	A robot writer for finance, housing, the worlds highlight and so on
	Search for Toutiao and Douyin

TikTok uses object detection algorithms, which currently uses AI solutions for fast and accurate detections, to identify several body parts of a person. It is an evolutionary force for social media, TikTok turns out to become a competitor to many other entertainment Apps, such as Facebook and Instagram.

The technology is also applied to the information extraction technology in the medical field to support the extraction of relevant factual basis information from the text of medical records and treatment history, and finally form a structured medical database that can be used by humans.

The information extraction and analysis technology of medical imaging is applied to construct a patient cohort. The patient group imaging diagnosis information is extracted and sub-constructed to construct a patient cohort, which is mainly reflected in epidemiological research, administrative management and so on. The extraction and analysis of individual patient imaging diagnosis information is correct Provide assistance in the treatment of individual diseases of patients; extraction and analysis of medical imaging process information for quality evaluation and improvement of medical imaging reports.

4. What can we do with NLP results?

The actual application of NLP results is based on the use of computers to process, understand and generate human languages (such as English, Chinese, and other minor languages, etc.) to achieve the effect of intelligent processing, and achieve the ability to exchange information between humans and machines for the purpose of communication.

Through the above analysis, we learn that NLP has constructed a series of results applications that can meet life scenarios based on a large and clear functional system. In the following section, we will introduce the practical application of the current NLP research results in human life from several the areas of business, media and medical care.

In the business field, chatbots used to capture potential customers are applied to major websites, such as Facebook and Intercom.

"The latest research shows that by entering a few lines to describe the desired characteristics, you can create a chatbot with a background story or even a personality appearance." (Kate Koidan, 2019) perfectly explains NLP's human-like interaction and empathy ability.

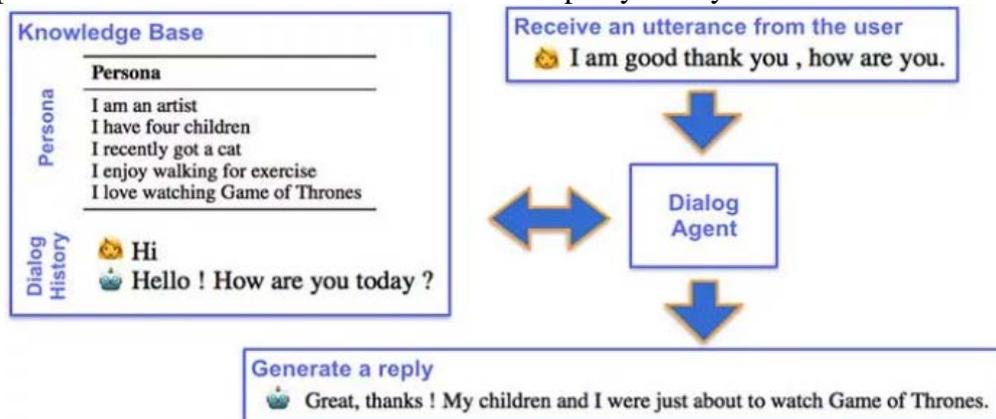


Figure. 3 Conversational AI with a persona (Scheme by HuggingFace)

In addition, chatbots realize user assistance through virtual services and provide standard solutions for common product problems, thereby reducing the actual consumption of customer service personnel and improving the efficiency of commodity transactions.

With the help of the flexible collaboration of CUI (Conversational User Interface) and NLP, our humans use voice for instructions, computers perform tasks after recognition, and the visualized results can interact with users in the future. Apple's assistant Siri is a task robot launched in October 2011 and is one of the most famous assistants. In many practices, task robots have become more flexible and commercialized than question answering robots. It can serve business in many common life scenarios, such as booking train tickets, plane tickets, movie tickets, bank card/credit card business processing, new retail, smart home, etc. Mission robots have now penetrated into business travel and finance.

NLP helps companies to target their core audiences by digitally positioning the market at the right time, and advertising to core consumer groups to increase conversion rates and close deals. Its constantly updated database is widely used in customer sentiment analysis and opinion mining in the business field, which can not only meet the needs of enterprises to obtain consumer suggestions and user experience, but also expand product functions and interactive experience, and enhance customer stickiness.

Now we have come to the era of new media, as users watch the short videos on new media platforms, they can identify whether they are in line with their own hobbies based on tag/category browsing, and the algorithm will recommend relevant contents to them according to their preferences and interests based on past browsing data. Content and products, or direct search by keywords in the search bar, all rely on the application of NLP technology.

NLP can also become an application technology that support for optimizing personalized recommendations for short videos. It can accurately understand human language logic and select personalized products from mass videos to push to users, creating convenience and providing more refined experiences for users.

NLP is widely used in product intelligence, including news recommendation, search, sensitive words, classification, etc. In video applications, it can achieve tasks including video understanding, video review, video labeling system, and short video production. At the same time, it can be used in content production. The automation and content review part can also provide auxiliary functions such as error correction, assessment scoring, and label optimization, giving more possibilities to the new media field under big data.

NLP not only has outstanding performance in media content construction, but also has made great breakthroughs in the field of media operation. The application of robot assistants to community operations can play a role in data growth. Robot assistants have data analysis and advertising monitoring, keyword reply, sign-in, intelligent companion chat and other functions, focusing on content dissemination and monetization technology, providing peace of mind services for knowledge-based communities.

In the medical field, the application of natural language processing in some research results has surpassed human labor efficiency and has become a recognized important tool for healthcare and life sciences. The data results reported by patients are very valuable to life science companies, who want to understand the safety of treatment products for a wide range of people in the long-term efficacy, and also provide raw materials for analysis for healthcare payers and medical systems.

Medical staff can use NLP to identify CHF risk factors from unstructured data (such as family medical history and smoking status) derived from continuous care documents, so that payers can more accurately collect important information and effectively assess population risks.

Among the existing achievements, the following successful cases can be used as reference data for the application of natural language processing research results in the medical field. Amazon has

successfully applied NLP to patient cohort analysis. The software assists doctors in analyzing patient data and drawing conclusions in advance, helping new drugs in clinical trials in the process of being admitted to find the correct patient, and finally finding the drug that best matches the patient's condition, which can improve the effectiveness of medical work and reduce the pressure on medical patients.

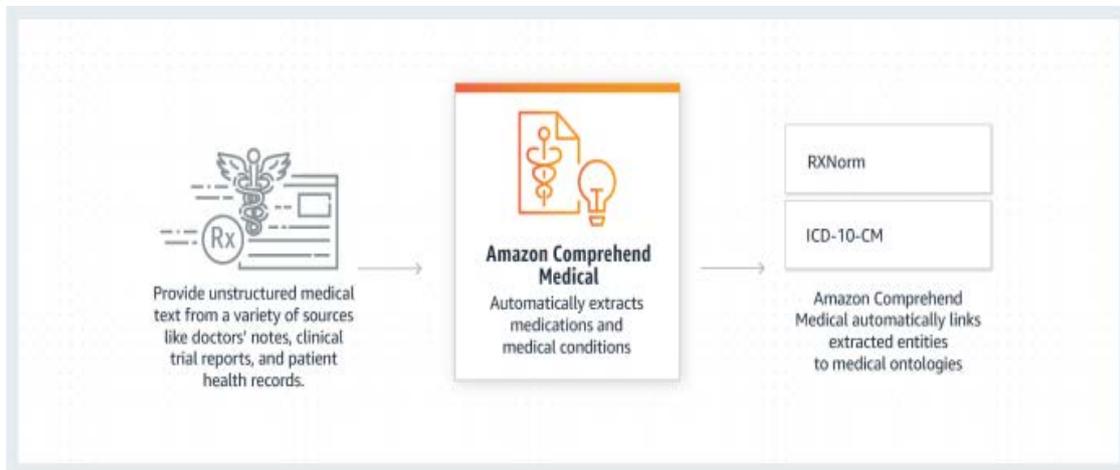


Figure. 4 Aws.amazon.com

3M (a multinational integrated manufacturing company that originated in the United States) invented a software called Code Ryte Code Assist System, which is used in natural language processing to scan medical applications that contain unstructured data to provide assistance, helping medical professionals and companies that provide health care products to collect the patient's disease status, treatment history, and surgery and services received, enrich the disease and patient database for the medical industry, just to make it intelligent and digital.

5. The dilemma needs to be solved

The research field of NLP has integrated the achievements of many disciplines in the exploration of this technology in the past few decades, it has basically become an important tool for communication between computer technology and human brain. But nowadays, in fact, artificial intelligence has appeared a bottleneck in the development of NLP applications.

At present, one of the main reasons why this research has run into a bottleneck is that there are hidden dangers in the setting of the principle. The process of realizing the mutual input, parsing and output of computer technology and natural language by artificial intelligence can be divided into the following steps: the hierarchical reception of words and sentences in the original language, the hierarchical transformation of words and sentences in the targeted language, and the generation of the final targeted language. In the early days, representative features were manually selected to combine and establish patterns to feed back to the computer for new matching, to the emergence of the concept of "deep learning" to realize computer autonomous learning and analysis from large-scale data. To a large extent, the problems such as data deviation are avoided, and clustering and merging operations in vector space are completed, but there are still some unavoidable problems.

First of all, in the process of hierarchical analysis of the original language, the computer acquiesced in a premise: human understanding of natural language needs to be the same hierarchical, and words can be separated and analyzed separately (Yin & Dong, 2008). In other words, while it is not yet certain whether human understanding of natural language is hierarchical, computers use

words and sentences hierarchically to parse the meaning of the original text, so as to simulate human brain neurons with computer component data, there are obvious loopholes in this concept.

Besides, one of the survival skills of NLP is to study the computing model through a large number of data statistics, so when new data appears, we can identify the data content through the recorded lexical features. Among them, Word Net hosted by George A. Miller and Frame Net project hosted by C. Fillmore are represented, which rely on a large number of semantic knowledge bases (Yin & Dong, 2008). However, the way to identify words and sentences through these semantic knowledge bases is difficult to identify context and articles, so it cannot achieve more accurate interaction and cannot have a more accurate long-paragraph text recognition ability. In a special context, there is a lack of partial analysis of human "emotion", that is, there are insurmountable obstacles in semantic analysis, so NLP cannot be truly "flawless"

Nowadays, artificial intelligence has been steadily upgraded in the fields of data analysis and statistics, which makes it handy in information analysis and processing in medical, financial and other fields. However, when chatbots and other fields need to face higher standards of humanoid interaction and empathy, it will be inadequate with the continuous updating of user requirements and the change of market demand. With the rapid development of technology, artificial intelligence users are not only satisfied with simple human-computer interactive chat, but also the enhancement of their emotional sustenance on the product is always a difficulty that artificial intelligence language processing needs to overcome.

6. The future prospect

First of all, academic integration, Mathematical Linguistics as the theoretical basis of NLP, Mathematical Linguistics combines language, computer science, mathematics and artificial intelligence, will continue to play an important role in all the fields of artificial intelligence in the future, and make language statistics more precise in future research. In addition, psychology, philosophy, cognitive science and other disciplines will be paid more and more attention in the field of natural language processing in the future, and the contents of these disciplines will form an electronic society within the computer, making the machine break through the Turing test to reach the state of "thinking". It also plays an important role in promoting the progress of artificial intelligence from morphological analysis to semantic analysis.

The second is the optimization of NLP model. As what has mentioned earlier, there are obvious flaws in the current natural language processing systems that rely on probabilistic and statistical models, and simple technical improvements can only reduce the loopholes but cannot fill it effectively. Therefore, it is necessary to propose a more sophisticated model concept to deal with the interactive needs in different contexts and large-scale real texts.

Computers are easier and more efficient in simply dealing with "translated words" because words themselves exist as simple data individuals as computers, while the real text, as the original language, is miserably mixed with the perceptual elements of the input. For a long time, scholars have debated the question of "whether perceptual components should be added to artificial intelligence in the process of development".

From the perspective of rational analysis mode, the replacement of human role by artificial intelligence is mainly reflected in that it can instantly complete thousands of calculations from huge data, which is incomparable to the human brain. However, the lack of sensibility of artificial intelligence also makes it impossible to achieve completing man-machine dialogue or form a proscription relationship in the fields of language, design, communication and so on. Therefore, in the current era when artificial intelligence has entered a new stage, how to balance the proportion of

perceptual elements in artificial intelligence and make a perfect combination of rational analysis and perceptual experience is still a topic that needs to be explored constantly.

7. Conclusion

Since the end of the 1940s, NLP has been widely used in medical, commercial, entertainment, communication and other fields from the beginning of the war to the present. With the changes of the times, it has brought many creative products to the society, and witnessed the rapid progress of science and technology and the change of people's needs. It has become an important existence in the applied language discipline. Although NLP still faces difficulties in concept, mode and ethics to overcome, its practicability and research ability make it a hot research topic in today's information network era. We can also foresee that in the near future, NLP will become an important treasure in the scientific community and occupy an important position in the needs of people's daily life.

Acknowledgements

Our deepest gratitude goes first and foremost to Professor Qu, Professor Zhang and Professor Lou from Harbin Institute of Technology, without their illuminating and consistent instruction, this paper could not have reached its present form. Their standards of academic excellence have made this journey a gratifying experience. Second, we would like to express our heartfelt gratitude to our friends and classmates who gave us their help and time in listening to us and helping us work out the problems during all the stages of writing this paper. Their brilliant ideas and perceptive observations have proved immensely constructive. Furthermore, we extend our thanks to our families for the encouragement that they gave us when we were working at this paper. Above all, none of this would have been possible without the help from them.

References

- [1] Carrell, D. S., Halgrim, S., & Tran, D., T. (2014). *Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence*. *American Journal of Epidemiology*, 179 (6): 749.
- [2] Dang, P. A., Kalra, K., & Blake, M., A. (2008). *Natural language processing using online analytic processing for assessing recommendations in radiology reports*. *Journal of the American College of Radiology*, 5 (3): 197-204.
- [3] Do, B. H., Wu, A., & Biswal, S. (2010). *Informatics in radiology: RADTF: a semantic search-enabled, natural language processor-generated radiology teaching file*. *Radiographics*, 30 (7): 2039-48.
- [4] Eggers, W. D., Mailk, N., & Gracie, M. (2019). *Using AI to unleash the power of unstructured government data*. Retrieved September 28, 2020, from <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html>
- [5] Feng Zhiwei. (2008). *History and present situation of natural language processing*. *Journal of Chinese Foreign Language*.
- [6] Garbade, D. (2018, October 15). *A Simple introduction to natural language processing*. Retrieved September 28, 2020, from <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32>
- [7] Gennaro Cuofano. (2020, August 11). *TikTok business model: the rise of creative social media powered by AI*. Retrieved September 28, 2020, from <https://fourweekmba.com/tiktok-business-model/>
- [8] Hou Zhixia & Cao Jun. (2003). *Development and prospect of natural language processing*. *Foreign language teaching in Shandong Province*.
- [9] Liddy, E. D. (2011). *Natural language processing*. In *Encyclopedia of Library and Information Science*, 2nd Ed.
- [10] Li Ying (2009). *Research on the extraction method of textual medical record information*. Hangzhou: Zhejiang University.
- [11] Paziienza, M. T. (1997). *International summer School on information extraction: a multidisciplinary approach to an emerging information technology*. Springer-Verlag, 1997: 425-426.

- [12] Sada Y, Hou J, Richardson P, (2013). *Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. Medical Care.*
- [13] Wang Ting, Mai Fanjin., & Liu Zhong. (2006). *Research on Natural language processing and its application prospect. Journal of Guilin College of Aerospace Industry.*
- [14] Wei Xiaoning. (2005). *Application of artificial intelligence in natural language understanding. Journal of Chinese Science and Technology Information.*
- [15] Yin Jin & Dong Jiarong. (2008). *The Development Trend of Natural Language Processing. The Research on Dialectics of Nature.*