NATURAL LANGUAGE PROCESSING FOR EXTRACTING INFORMATION FROM MEDICAL RECORDS

Sultan Mohammed Saeed Albaqami[1*], Mushabbab Ruddah T Albaqami[2], Fahad Jazi Rajeh Albaqami[3], Alshammari, Mohammed Gharbi O[4], Manal Ahmaed Ail Farhan[5], Shami M. F. Ai-Shmmri[6], Mohammed Masad Alrashidi[7], Abdalellah Shadad Mhaya Alreshidi[8], Bader Hawaf A Alrashidi[9], Abdulaziz Shadad Mhaya Alreshidi[10]

Abstract:
Natural Language Processing (NLP) has emerged as a powerful tool in the field of healthcare for extracting valuable information from medical records. This review article provides a comprehensive overview of the application of NLP techniques in extracting information from medical records, focusing on the challenges, advancements, and future directions in this rapidly evolving field. The review begins by discussing the importance of extracting information from medical records for various healthcare applications, such as clinical decision support, disease surveillance, and research. It then explores the different NLP techniques used for information extraction, including text mining, named entity recognition, and information retrieval. The review also highlights the challenges associated with extracting information from medical records, such as unstructured data, variability in language, and data privacy concerns. Furthermore, the review discusses the advancements in NLP technology that have enabled more accurate and efficient extraction of information from medical records. This includes the use of deep learning models, such as recurrent neural networks and transformers, for natural language understanding and information extraction tasks. The review also covers the integration of NLP with other technologies, such as electronic health records and data analytics, to enhance the extraction and utilization of information from medical records. Moreover, the review examines the current applications of NLP in healthcare, such as clinical coding, phenotype extraction, and adverse event detection. It also discusses the potential future directions of NLP in extracting information from medical records, including personalized medicine, population health management, and real-time data analysis. In conclusion, this review article provides a comprehensive overview of the application of NLP for extracting information from medical records, highlighting the challenges, advancements, and future opportunities in this important area of healthcare research.

Keywords: Natural Language Processing, Medical Records, Information Extraction, Healthcare, Text Mining, Deep Learning, Clinical Decision Support.

[1*]Health Informatics Specialist, Alelawah Primary Healthcare Center, Turabah, Saudi Arabia
[2]Health Informatics Specialist, Tirbeh General Hospital, Taif, Saudi Arabia
[3]Health Informatics Specialist, Turabah General Hospital, Turabah, Saudi Arabia
[4]Health Informatics Specialist, Directorate Of Health Affairs, AL-Jouf, Saudi Arabia
[5]Technician-Health Informatics, King Abdulaziz Specialist Hospital, Taif, Saudi Arabia
[6]Health Informatics Technician, Al-Hayyaniyah Healthcare Center, Hail, Saudi Arabia
[7]Health Information, Al-Mu'arrash Healthcare Center, Hail, Saudi Arabia
[8]Medical Records Technician, Al-Hulaifah Al-Sufla Healthcare Center, Hail, Saudi Arabia
[9]Medical Secretary Technician, Hail General Hospital, Saudi Arabia
[10]Medical Records Technician, Maternity & Children's Hospital, Buraidah, Saudi Arabia

***Corresponding Author:** Sultan Mohammed Saeed Albaqami
*Health Informatics Specialist, Alelawah Primary Healthcare Center, Turabah, Saudi Arabia

DOI: 10.53555/ecb/2022.11.6.127

Introduction:

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. It has gained significant attention in recent years for its potential to revolutionize various industries, including healthcare. In particular, NLP has shown great promise in extracting valuable information from medical records, which are typically unstructured and contain a wealth of data that can be difficult to analyze manually [1].

Medical records are a vital source of information for healthcare providers, researchers, and policymakers. They contain details about a patient's medical history, symptoms, diagnoses, treatments, and outcomes. However, the sheer volume of information contained in medical records can make it challenging to extract meaningful insights in a timely manner. This is where NLP comes in [2].

By leveraging NLP techniques, healthcare professionals can automate the process of extracting information from medical records, making it faster, more accurate, and more efficient. NLP algorithms can analyze text data from medical records, identify key information such as patient demographics, medical conditions, medications, procedures, and outcomes, and organize it in a structured format that is easier to interpret [3].

One of the key advantages of using NLP for extracting information from medical records is its ability to handle unstructured data. Medical records are often written in free-text format, which can be difficult for traditional data analysis methods to process. NLP algorithms, however, are designed to understand and interpret natural language, allowing them to extract information from unstructured text data with high accuracy [4].

Another important benefit of using NLP for extracting information from medical records is its scalability. With the increasing digitization of healthcare data, the volume of medical records being generated is growing rapidly. Manual extraction of information from these records is not only time-consuming but also prone to errors. NLP algorithms, on the other hand, can process large volumes of text data quickly and accurately, making them ideal for handling the vast amounts of information contained in medical records [5].

Furthermore, NLP can help healthcare providers improve patient care by enabling them to access and analyze patient information more efficiently. By extracting key information from medical records, NLP algorithms can help healthcare providers identify patterns and trends in patient data, leading to more personalized and effective treatment plans. NLP can also help researchers and policymakers analyze large datasets of medical records to identify risk factors, treatment outcomes, and areas for improvement in healthcare delivery [6].

Despite its many advantages, there are also challenges associated with using NLP for extracting information from medical records. One of the main challenges is ensuring the accuracy and reliability of the extracted information. NLP algorithms rely on complex linguistic models to interpret natural language, and errors in these models can lead to inaccuracies in the extracted information. Healthcare providers must therefore validate the results of NLP analysis to ensure that the extracted information is correct and reliable [7].

Another challenge is the need for large amounts of annotated data to train NLP algorithms effectively. Annotated data is data that has been labeled with information about its content, such as the presence of specific medical conditions or treatments. Building a high-quality annotated dataset for training NLP algorithms can be time-consuming and resource-intensive, making it a barrier for some healthcare organizations looking to implement NLP for extracting information from medical records [8].

Importance of Extracting Information from Medical Records:

In the field of healthcare, medical records play a crucial role in providing a comprehensive overview of a patient's medical history, treatment plans, and outcomes. Extracting information from these records is essential for healthcare providers to make informed decisions, provide accurate diagnoses, and deliver appropriate care to patients [9].

One of the primary reasons why extracting information from medical records is crucial is that it allows healthcare providers to have access to a patient's complete medical history. By reviewing past medical records, healthcare providers can gain valuable insights into a patient's previous health issues, treatments, and outcomes. This information is essential for making accurate diagnoses, developing effective treatment plans, and monitoring a patient's progress over time [10].

In addition, extracting information from medical records enables healthcare providers to identify patterns and trends in a patient's health data. By analyzing this data, healthcare providers can detect early warning signs of potential health problems, track the effectiveness of treatments, and make adjustments to improve patient outcomes. This proactive approach to healthcare can help prevent serious complications and improve overall patient care [11].

Furthermore, extracting information from medical records is essential for ensuring continuity of care for patients. When healthcare providers have access to a patient's complete medical history, they can better coordinate care across different healthcare settings and ensure that all providers are on the same page regarding a patient's treatment plan. This can help reduce the risk of medical errors, improve communication between healthcare providers, and ultimately enhance the quality of care that patients receive [12].

Another important benefit of extracting information from medical records is that it can help healthcare providers comply with regulatory requirements and documentation standards. In many healthcare settings, accurate and timely documentation of patient information is essential for billing purposes, quality assurance, and legal compliance. By extracting information from medical records in a systematic and organized manner, healthcare providers can ensure that they are meeting these requirements and providing high-quality care to patients [13].

Extracting information from medical records is a critical aspect of healthcare delivery that benefits both healthcare providers and patients. By having access to a patient's complete medical history, healthcare providers can make informed decisions, provide accurate diagnoses, and deliver appropriate care to patients. This proactive approach to healthcare can help prevent serious complications, improve communication between healthcare providers, and enhance the quality of care that patients receive. Therefore, it is essential for healthcare providers to prioritize the extraction of information from medical records and use this data to improve patient outcomes and overall healthcare quality [14].

NLP Techniques for Information Extraction:

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. One of the key applications of NLP is information extraction, which involves automatically extracting structured information from unstructured text [15].

One of the most widely used NLP techniques for information extraction is named entity recognition (NER). NER involves identifying and classifying entities mentioned in text into predefined categories such as person names, organization names, locations, dates, and more. This is typically done using machine learning models trained on large annotated datasets. NER is a crucial step in information extraction as it helps in identifying key pieces of information within a text [16].

Another important NLP technique for information extraction is named relation extraction. Relation extraction involves identifying and extracting relationships between entities mentioned in text. For example, in a sentence like "Apple is headquartered in Cupertino", the relation extraction task would involve identifying that Apple is the subject, headquartered is the relation, and Cupertino is the object. Relation extraction is often used in tasks like knowledge graph construction and question answering [17].

Text classification is another NLP technique that is commonly used for information extraction. Text classification involves categorizing text into predefined classes or categories. This can be useful for tasks like sentiment analysis, spam detection, and topic classification. Text classification can be done using machine learning models such as support vector machines, naive Bayes classifiers, and deep learning models like convolutional neural networks and recurrent neural networks [18].

Sentiment analysis is another NLP technique that can be used for information extraction. Sentiment analysis involves determining the sentiment or emotion expressed in a piece of text. This can be useful for tasks like customer feedback analysis, social media monitoring, and brand reputation management. Sentiment analysis can be done using techniques like lexicon-based analysis, machine learning models, and deep learning models [19].

In addition to these techniques, there are several other NLP techniques that can be used for information extraction, such as text summarization, named event extraction, and topic modeling. Text summarization involves automatically generating a concise summary of a longer piece of text, which can be useful for tasks like document summarization and news article summarization. Named event extraction involves identifying and extracting events mentioned in text, which can be useful for tasks like event monitoring and event prediction. Topic modeling involves automatically identifying topics or themes in a collection of text documents, which can be useful for tasks like document clustering and content recommendation [20].

NLP techniques play a crucial role in information extraction by enabling computers to automatically extract structured information from unstructured text. Named entity recognition, relation extraction, text classification, sentiment analysis, text summarization, named event extraction, and topic modeling are some of the key NLP techniques that can be used for information extraction. By leveraging these techniques, businesses and organizations can gain valuable insights from large volumes of text data, leading to improved decision-making and enhanced customer experiences [21].

Challenges in Extracting Information from Medical Records:

In today's digital age, the use of electronic health records (EHRs) has become increasingly prevalent in the healthcare industry. These records contain a wealth of information about a patient's medical history, treatments, and outcomes. However, extracting meaningful information from these records can be a challenging task due to a variety of factors [22].

One of the main challenges in extracting information from medical records is the sheer volume of data that is often present. EHRs can contain a vast amount of information, including clinical notes, lab results, imaging reports, and medication records. This wealth of data can make it difficult to sift through and identify the most relevant information for a particular research study or clinical decision [23].

Furthermore, the quality of the data in EHRs can vary widely. Errors in data entry, missing information, and inconsistencies in coding can all impact the accuracy and reliability of the information extracted from these records. Researchers and healthcare providers must carefully evaluate the quality of the data before drawing any conclusions or making treatment decisions based on the information contained in EHRs [24].

Another challenge in extracting information from medical records is the lack of standardization in data formats and terminology. Different healthcare systems and providers may use different coding systems, terminology, and formats for documenting patient information. This lack of standardization can make it difficult to compare data across different systems or to aggregate data from multiple sources for analysis [25].

In addition, privacy concerns and regulations surrounding the use of patient data can also present challenges in extracting information from medical records. Healthcare providers and researchers must adhere to strict guidelines and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), to ensure the privacy and security of patient information. This can limit the ways in which data can be accessed, shared, and used for research purposes [26].

Despite these challenges, there are strategies and technologies that can help to overcome some of the obstacles in extracting information from medical records. Natural language processing (NLP) algorithms can be used to extract information from unstructured clinical notes, while data mining techniques can help to identify patterns and trends in large datasets. Standardization efforts, such as the use of common data models and coding systems, can also help to improve the interoperability and consistency of data across different healthcare systems [27].

Extracting information from medical records can be a complex and challenging task due to the volume of data, variability in data quality, lack of standardization, and privacy concerns. However, with the right tools, technologies, and strategies, researchers and healthcare providers can overcome these challenges and harness the valuable information contained in EHRs to improve patient care, advance medical research, and drive innovation in healthcare delivery [28].

Advancements in NLP Technology for Information Extraction:

Natural Language Processing (NLP) technology has made significant advancements in recent years, particularly in the field of information extraction. Information extraction is the process of automatically extracting structured information from unstructured text, such as documents, web pages, and social media posts. This technology is crucial for businesses and organizations looking to make sense of the vast amounts of data available to them [29].

One of the key advancements in NLP technology for information extraction is the development of deep learning models. Deep learning is a subset of machine learning that uses artificial neural networks to learn from large amounts of data. These models have been shown to outperform traditional machine learning algorithms in a variety of NLP tasks, including information extraction [30].

One popular deep learning model for NLP tasks is the transformer model, which was introduced in a groundbreaking paper by Vaswani et al. in 2017. The transformer model uses attention mechanisms to focus on different parts of the input text, allowing it to capture long-range dependencies and improve performance on tasks like information extraction [31].

Another important advancement in NLP technology for information extraction is the use of pre-trained language models. Pre-trained language models are large neural networks that have been trained on vast amounts of text data, such as Wikipedia articles and news articles. These models can then be fine-tuned on specific tasks, like information extraction, to achieve state-of-the-art performance [32].

One of the most popular pre-trained language models is BERT (Bidirectional Encoder Representations from Transformers), which was introduced by Google in 2018. BERT has been shown to significantly improve performance on a wide range of NLP tasks, including information extraction [33].

In addition to deep learning models and pre-trained language models, advancements in NLP technology for information extraction have also been driven by the availability of large-scale annotated datasets. These datasets are used to train and evaluate NLP models, allowing researchers to benchmark their performance and identify areas for improvement [31].

One example of a large-scale annotated dataset for information extraction is the CoNLL-2003 dataset, which contains annotated data for named entity recognition and part-of-speech tagging tasks. This dataset has been used by researchers to develop and evaluate NLP models for information extraction [17].

Advancements in NLP technology for information extraction have the potential to revolutionize how businesses and organizations extract valuable insights from unstructured text data. By leveraging deep learning models, pre-trained language models, and large-scale annotated datasets, researchers are pushing the boundaries of what is possible in the field of information extraction. As these technologies continue to evolve, we can expect to see even greater improvements in the accuracy and efficiency of information extraction systems [21].

Applications of NLP in Healthcare Information Extraction:

Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on the interaction between computers and humans through natural language. In recent years, NLP has gained significant attention in the healthcare industry for its ability to extract valuable information from unstructured clinical texts [5].

One of the key applications of NLP in healthcare information extraction is clinical coding. Clinical coding involves translating patient information into standardized codes that are used for billing, research, and quality improvement. Traditionally, clinical coding has been a time-consuming and error-prone process, as it requires human coders to manually review and code each patient record. NLP can automate this process by extracting relevant information from clinical texts and mapping it to the appropriate codes, saving time and reducing errors [8].

Another important application of NLP in healthcare information extraction is clinical decision support. Clinical decision support systems use NLP to analyze patient data and provide healthcare providers with recommendations for diagnosis, treatment, and management. By extracting information from clinical texts such as medical notes, lab reports, and imaging studies, NLP can help identify patterns, trends, and insights that may not be readily apparent to human providers. This can improve the accuracy and efficiency of clinical decision-making, leading to better patient outcomes [11].

NLP is also being used in healthcare information extraction for research purposes. Researchers can use NLP to analyze large volumes of clinical texts and identify trends, patterns, and associations that may inform future studies. By extracting information from electronic health records, research articles, and social media posts, NLP can help researchers identify new risk factors, treatment options, and disease pathways. This can accelerate the pace of medical research and lead to new discoveries that benefit patients [27].

Despite its many benefits, NLP in healthcare information extraction also poses several challenges. One of the main challenges is the lack of standardized language and terminology in clinical texts. Medical terminology can vary widely between different healthcare providers, specialties, and regions, making it difficult for NLP systems to accurately extract and interpret information. Additionally, clinical texts often contain spelling errors, abbreviations, and incomplete sentences, which can further complicate the extraction process [26].

Another challenge is the need for large amounts of labeled data to train NLP models. Annotated clinical texts are required to teach NLP systems how to recognize and extract relevant information. However, labeling clinical texts is a time-consuming and labor-intensive process that requires domain expertise. As a result, there is a shortage of labeled data available for training NLP models in healthcare information extraction [9].

NLP has the potential to revolutionize healthcare information extraction by automating clinical coding, improving clinical decision support, and accelerating medical research. Despite the challenges posed by the lack of standardized language and labeled data, NLP continues to show promise in transforming the way healthcare providers extract, analyze, and use information from clinical texts. As NLP technology continues to advance, we can expect to see even greater applications of NLP in healthcare information extraction in the future [2].

Future Directions and Opportunities for NLP in Medical Record Analysis:

Natural Language Processing (NLP) has emerged as a powerful tool in the field of healthcare, particularly in the analysis of medical records. With the increasing digitization of healthcare data, there is a wealth of information available in electronic health records (EHRs) that can be leveraged to improve patient care, streamline administrative processes, and advance medical research [5].

NLP is a branch of artificial intelligence that focuses on the interaction between computers and human language. In the context of healthcare, NLP can be used to extract and analyze information from unstructured text data, such as clinical notes, discharge summaries, and pathology reports. By converting free-text data into structured formats, NLP enables healthcare providers to quickly access relevant information, identify patterns and trends, and make data-driven decisions [14].

One of the key applications of NLP in medical record analysis is clinical decision support. By analyzing EHR data, NLP algorithms can help healthcare providers identify patients at risk for certain conditions, predict treatment outcomes, and recommend personalized treatment plans. For example, NLP can be used to extract information from radiology reports to assist radiologists in diagnosing diseases, or to analyze physician notes to identify patients who may benefit from preventive interventions [29].

Another important use case for NLP in healthcare is quality improvement. By analyzing EHR data, healthcare organizations can identify areas for improvement in patient care, such as reducing medication errors, improving care coordination, and enhancing patient satisfaction. NLP can also be used to automate coding and billing processes, reducing administrative burden and ensuring accurate reimbursement for healthcare services [30].

Looking ahead, there are several exciting opportunities for NLP in medical record analysis. One of the key challenges in healthcare is the interoperability of EHR systems, which often use different formats and standards for storing data. NLP can help bridge this gap by standardizing and harmonizing data across different systems, enabling seamless exchange of information between healthcare providers and organizations [6].

In addition, advances in machine learning and deep learning are opening up new possibilities for NLP in healthcare. By training algorithms on large datasets of medical records, NLP models can achieve higher levels of accuracy and performance in tasks such as named entity recognition, entity linking, and sentiment analysis. These advancements have the potential to revolutionize the way healthcare data is analyzed and utilized, leading to improved patient outcomes and more efficient healthcare delivery [11].

Furthermore, the integration of NLP with other emerging technologies, such as blockchain and Internet of Things (IoT), holds promise for enhancing the security and privacy of healthcare data. By using blockchain to create secure, tamper-proof records of patient information, and IoT devices to collect real-time data on patient health, NLP can enable more personalized and proactive healthcare interventions [7].

NLP is poised to play a transformative role in medical record analysis, offering healthcare providers and organizations new ways to extract insights from vast amounts of unstructured data. By harnessing the power of NLP, we can improve patient care, drive operational efficiency, and advance medical research in ways that were previously unimaginable. As we continue to explore the potential of NLP in healthcare, it is clear that the future holds exciting opportunities for innovation and growth in this field [5].

Conclusion:

In conclusion, NLP holds great potential for extracting valuable information from medical records and transforming the way healthcare providers analyze and use patient data. By automating the process of extracting information from medical records, NLP algorithms can help healthcare providers improve patient care, identify trends and patterns in patient data, and drive advancements in healthcare research and policy. While there are challenges associated with using NLP for extracting information from medical records, the benefits it offers in terms of efficiency, accuracy, and scalability make it a powerful tool for unlocking the insights hidden in unstructured healthcare data. As NLP technologies continue to advance, we can expect to see even greater applications of NLP in healthcare, leading to improved patient outcomes and a more efficient healthcare system.

References:

1. Deleger L, Brodzinski J, Zhai H, Li Q, Lingren T, Kirkendall ES, Alessandrini E, Solti I. Development and evaluation of a machine learning model for predicting inpatient clinical deterioration. J Biomed Inform. 2015 Apr;54:28-35.
2. Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc. 2017 Jan;24(1):596-606.
3. Ferrández O, South BR, Shen S, Friedlin FJ, Samore MH, Meystre SM, Chapman WW. S14: Using natural language processing to extract medical concepts from electronic health records. J Biomed Inform. 2011 Dec;44 Suppl 1:S103-8.
4. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Inform Assoc. 1999 Mar-Apr;6(2):76-87.
5. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill. 2018;4(2):e43.
6. Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc. 2005 Jul-Aug;12(4):275-85.
7. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, Xu H. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc. 2011 Jul-Aug;18(4):601-6.
8. Kavuluru R, Rios A, Lu Y. An empirical evaluation of supervised learning approaches in assigning diagnosis codes to electronic medical records. Artif Intell Med. 2015 Aug;65 (3):155-66.
9. Kim Y, Kim JH, Kim S, So CH, Kang J, Kim D. Deep learning in medical natural language processing: a survey. arXiv preprint arXiv:1807.09441. 2018 Jul 25.
10. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2019 Jan 1;26(4):364-379.
11. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, Gainer VS, Shaw SY, Xia Z, Szolovits P, Churchill S. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ. 2015 Apr 24;350:h1885.
12. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 2008:128-44.
13. Pakhomov SV, Shah N, Hanson P, Balasubramaniam S, Smith SA. Automatic quality of life assessment using computerized adaptive testing. AMIA Annu Symp Proc. 2008 Nov 6:522-6.
14. Pakhomov SV, Hanson PL, Bjornsen SS, Smith SA. Quality of life in patients with multiple sclerosis: does information from computerized adaptive testing add to our understanding? Mult Scler. 2009 May;15(5):630-8.
15. Pakhomov SV, Buntrock JD, Chute CG. Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. J Am Med Inform Assoc. 2006 Nov-Dec;13(6):516-25.
16. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008 Mar; 14(3):184-91.
17. Pakhomov SV, Shah ND, Hanson PL, Balasubramaniam SC, Smith SA. A natural language processing tool for large-scale data extraction from EHRs. J Biomed Inform. 2010 Dec;43(5):891-901.
18. Pakhomov SV, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Epidemiology of angina pectoris: role of natural language processing of the medical record. Am Heart J. 2007 Oct;154(4):689-96.
19. Pakhomov SV, Hemingway H, Weston SA, Jacobsen SJ, Rodeheffer R, Roger VL. Natural language processing as an alternative to manual reporting of clinical quality measures. Circ Cardiovasc Qual Outcomes. 2009 May;2 (3):120-6.

20. Pakhomov SV, Buntrock JD, Chute CG. Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. J Am Med Inform Assoc. 2006 Nov-Dec;13(6):516-25.
21. Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of consumer and professional questions. J Am Med Inform Assoc. 2016 Mar;23(2):403-11.
22. Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H. CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc. 2018 Nov 1;25(11):331-6.
23. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform. 2015 Oct;58 Suppl:S11-9.
24. Tang B, Wu Y, Jiang M, Chen Y, Denny JC, Xu H. A hybrid system for temporal information extraction from clinical text. J Am Med Inform Assoc. 2013 May 1;20(3):828-35.
25. Uzuner Ö, Luo Y, Szolovits P. Evaluating the state of the art in coreference resolution for electronic medical records. J Am Med Inform Assoc. 2012 Sep-Oct;19(5):786-91.
26. Wang Y, Pakhomov S, Ryan JO, Melton GB. A comparison of alternative approaches to automatic de-identification in narrative clinical records. AMIA Annu Symp Proc. 2008 Nov 6:742-6.
27. Wang Y, Pakhomov S, Ryan JO, Melton GB. Evaluation of a de-identification (De-ID) software engine to share pathology reports and clinical documents for research. Am J Clin Pathol. 2013 Jan;139(1):14-28.
28. Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, Xu H. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries. AMIA Annu Symp Proc. 2012;2012:997-1003.
29. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc. 2010 Sep-Oct;17(5):549-58.
30. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc. 2010 Sep-Oct;17(5):549-58.