Natural Language Processing: A Literature Survey

Submitted in partial fulfilment of the requirements for the degree of

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

Sridhar Suraj

17BCE2245

Under the guidance of

Dr. SHARMILA BANU K

Associate Professor Grade 1

Department of Software Systems

School of Computer Science and Engineering

Vellore Institute of Technology

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

DECLARATION

I hereby declare that the thesis entitled “Natural Language Processing: A Literature Survey” submitted by me, for the award of the degree of Bachelor of Technology in Programme to VIT is a record of Bonafide work carried out by us under the supervision of Dr Sharmila Bhanu K.

I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place : Vellore

Date :

Signature of the Candidate

CERTIFICATE

This is to certify that the project entitled “Natural Language Processing: A Literature Survey” submitted by Sridhar Suraj, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, for the award of the degree of Bachelor of Technology in Computer Science and Engineering, is a record of bonafide work carried out by him under my supervision during the period, 2.09. 2020 to 16.10.2020, as per the VIT code of academic and research ethics.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The project fulfils the requirements and regulations of the University and in my opinion meets the necessary standards for submission.

Place : Vellore

Date :

Signature of the Candidate

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to our faculty in-charge Dr Sharmila Bhanu K as well as our Head of Department Prof Santhi.V who gave me a golden opportunity to do this wonderful project on the topic of Natural Language Process – a literature survey. This also helped me in doing a lot of research and learn many new things. I am really indebted to them.

Sridhar Suraj

17BCE2245

TABLE OF CONTENTS

CHAPTER -1

1.1 INTRODUCTION

Natural Language Processing is a very vast and versatile tool that is used for the manipulation of natural language, such as speech, text or by software. The study of Natural Language Processing has been around for 50 years and an ever growing field of computer science and technology. It is a fundamental element of Artificial Intelligence.

1.2 OBJECTIVES

- Paraphrase the input text.
- Sentiment Analysis and Opinion Mining
- Language Understanding
- Machine Translation
- Content Generation
- Main objective of the Literature Survey is to understand the applications of Natural Language Processing in various fields and domains.

1.3 SCOPE OF STUDY

- **Data Collection:**

  The data for this literature survey was collected and analyzed by going through multiple papers in the field of Natural Language Processing. The papers were collected from various fields such as Medical (Radiology, Oncology, Orthopedic) and Computer Science(Neural Networks, Social Media Security and Management) and Meteorology(Weather Forecasting)

- **Application of NLP in Medical Field:**

  i) Handle the Surge in Clinical Data:

  The increased use of patient health record systems and the digital transformation of medicine has led to a spike in the volume of data available with healthcare organizations. The need to make sense out of this data and draw credible insights happens to be a significant driver.

ii) Improving Clinical Documentation:

Electronic Health Record solutions often have a complex structure, so that documenting data in them is a hassle. With speech-to-text dictation, data can be automatically captured at the point of care, freeing up physicians from the tedious task of documenting care delivery.

iii) Empower patients with Health Literacy:

With conversational AI already being a success within the healthcare space, a key use-case and benefit of implementing this technology is the ability to help patients understand their symptoms and gain more knowledge about their conditions. By becoming more aware of their health conditions, patients can make informed decisions and keep their health on track by interacting with an intelligent chatbot.

iv) Assess the need for Higher Quality Health Care:

NLP can be the front-runner in assessing and improving the quality of healthcare by measuring physician performance and identifying gaps in care delivery.

Research has shown that artificial intelligence in healthcare can ease the process of physician assessment and automate patient diagnosis, reducing the time and human effort needed in carrying out routine tasks such as patient diagnosis. NLP in healthcare can also identify and mitigate potential errors in care delivery. A study showed that NLP could also be utilized in measuring the quality of healthcare and monitor adherence to clinical guidelines.

v) Identify patients who need improved healthcare:

Machine Learning and NLP tools have the capabilities needed to detect patients with complex health conditions who have a history of mental health or substance abuse and need improved care. Factors such as food insecurity and housing instability can deter the treatment protocols, thereby compelling these patients to incur more costs in their lifetime.

- **Applications of NLP in Computer and Technology:**

  i) Sentiment Analysis:

  Mostly used on the web & social media monitoring, Natural Language Processing is a great tool to comprehend and analyse the responses to the business messages published on social media platforms. It helps to analyse the attitude and emotional state of the writer (person commenting/engaging with posts). This application is also known as opinion mining. It is implemented through a combination of Natural Language Processing and statistics by assigning values to the text (positive, negative or neutral) and in turn making efforts to identify the underlying mood of the context (happy, sad, angry, annoyed, etc.).This application of NLP helps business organisations gain insights on consumers and do a competitive comparison and make necessary adjustments in business strategies, whenever required. Such data is also useful in designing a better customer experience and enhancing the product. Furthermore, sentiment analysis or emotion exploration is a great way to know about brand perception.

  ii) Chatbots:

  We hear a lot about Chatbots these days, chatbots are the solution for consumer frustration regarding customer care call assistance. They provide modern-day virtual assistance for simple problems of the customer and offload low-priority, high turnover tasks which require no skill. Intelligent Chatbots are going to offer personalised assistance to the customer in the near future.A lot of Industry analysts predict that Chatbots are an emergent trend which will offer real-time solutions for simple customer service problems. They are unquestionably gaining a lot of trust and popularity from the consumer as well as engineers. They are useful in providing standard solutions to common problems. Chatbots help save time, human efforts, cost and provide efficient solutions (and keep improving from learning) from time to time.

  iii) Market Intelligence:

  Business markets are influenced and impacted by market knowledge and information exchange between various organisations, stakeholders, governments and regulatory bodies. It is vital to stay up to date with industry

trends and changing standards. NLP is a useful technology to track and monitor the market intelligence reports for and extract the necessary information for businesses to build new strategies. Widely used in financial marketing, NLP gives exhaustive insights into employment changes and status of the market, tender delays, and closings, or extracting information from large repositories. These are some of the few applications of Natural Language Processing which will be witnessed by business organisations in the time to come. There are other applications as well, such as reputation monitoring, neural machine translation, hiring tools and management, regulatory compliance, data visualisation, biometrics, robotics, process automation etc. NLP is the key to the quest for general artificial intelligence since language is a key indicator of intelligence in our society.

CHAPTER-2

2.1 LITERATURE REVIEW

1) Natural Language Processing in Oncology

Importance:

Natural language processing (NLP) has the potential to accelerate translation of cancer treatments from the laboratory to the clinic and will be a powerful tool in the era of personalized medicine. This technology can harvest important clinical variables trapped in the free-text narratives within electronic medical records.

Observations:

Natural language processing can be used as a tool for oncological evidence-based research and quality improvement. Oncologists interested in applying NLP for clinical research can play pivotal roles in building NLP systems and, in doing so, contribute to both oncological and clinical NLP research. Herein, the authors provide an introduction to NLP and its potential applications in oncology, a description of specific tools available, and a review on the state of the current technology with respect to cancer case identification, staging, and outcomes quantification.

Conclusions and Relevance:

More automated means of leveraging unstructured data from daily clinical practice is crucial as therapeutic options and access to individual-level health information increase. Research-minded oncologists may push the avenues of evidence-based research by taking advantage of the new technologies available with clinical NLP. As continued progress is made with applying NLP toward oncological research, incremental gains will lead to large impacts, building a cost-effective infrastructure for advancing cancer care.

2) **Machine Learning Approach to Fake News Detection Using Knowledge Verification and Natural Language Processing**

Abstract:

The term “fake news” gained international popularity as a result of the 2016 US presidential election campaign. It is related to the practice of spreading false and/or misleading information in order to influence popular opinion. This practice is known as disinformation. It is one of the main weapons used in information warfare, which is listed as an emerging cybersecurity threat. In this paper, the authors explore “fake news” as a disinformation tool. The authors survey previous efforts in defining and automating the detection process of “fake news”. The authors establish a new fluid definition of “fake news” in terms of relative bias and factual accuracy. The authors devise a novel framework for fake news detection, based on our proposed definition and using a machine learning model.

Features:

The 5 NLP features are: stopwords percentage; ratio of proper nouns to nouns; title length; ARI readability; overall sentiment of the text using Google NLP API for sentiment analysis.

Conclusion:

In this paper the authors define fake news in the context of information warfare. The authors briefly study the socio-political implications of fake news and the authors investigate previous efforts in automating fake news detection. The authors find that the most promising framework for fake news detection uses a combination of source and fact verification and NLP analysis, and The authors propose a hybrid framework based on our previous work in automating incident classification.

3) **Detecting Phishing Attacks Using Natural Language Processing and Machine Learning:**

Abstract:

Phishing attacks are one of the most common and least defended security threats today. The authors present an approach which uses natural language processing techniques to analyze text and detect inappropriate statements which are indicative of phishing attacks. Our approach is novel compared to previous work because it focuses on the natural language text contained in the attack, performing semantic analysis of the text to detect malicious intent. To demonstrate the effectiveness of our approach, The authors have evaluated it using a large benchmark set of phishing emails.

Detection Algorithm:

Sea Hound:

1. define SEAHound(text)

2. bad, urgent, generic = False

3. foreach sentence s in text

4. bad | = BadQuestion(s) OR BadCommand(s)

5. urgent | = UrgentTone(s)

6. generic | = GenericGreeting(s)

7. link = LinkAnalysis(s)

8. if link

9. return True

10. if majority(bad, urgent, generic)

11. return True

12. return False

Conclusion:

The authors present an approach to detect targeted phishing email attacks. Our approach relies on analysis of the text, rather than metadata which might be associated with emails. As a result, our approach is effective for detecting phishing emails which are composed of pure text. Our results on phishing emails demonstrate significantly improved recall which demonstrates that semantic information is a strong indicator of social engineering.

CHAPTER-3

3. RESULTS AND ANALYSIS

After the analysis of 50 papers in the field of Natural Language Processing, we can infer the various uses of NLP techniques such as POS tagging, Tokenizing, Morphology, Segmentation and many other techniques. We have also found the various methods used in Medical NLP as well as Technical NLP.

NLP techniques used in medical field are:

Brok,

cTAKES,

LEXIMER,

MALLET

,RadLex®,

SAPHIRE,

YTEX

Rochester Epidemiology Project (REP)

NLP techniques used in Technology field:

Latent Semantic Index[LSI]

Bayesian Learning

Recognizing Textual Entitlement[RTE]

Hidden Markov Models[HMM]

Named entity recognition[NER]

Hermes Information Extraction Language (HIEL)

Relationships identification

Natural Language Generation(NLG)

CHAPTER-4 – SUMMARY AND CONCLUSION

Name & Reg.Number	Project Title	Project Description
SRIDHAR SURAJ - 17BCE2245	Literature Survey of Natural Language Processing(50 Papers)	Analysis of various research articles in the field of NLP.

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
1	Neural Transfer learning using NLP	Ruder, Sebastian	Canonical correlation analysis	http://hdl.handle.net/10379/15463,Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
2	Survey of Hate Speech Detection using NLP	Anna Schmidt, MIchael Wiegand	Support vector machines	Njagi Dennis Gitari, Zhang Zuping, Hanyurwimfura Damien, and Jun Long. 2015. A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering, 10(4):215–230
3	Recent Trends in Deep Learning Based Natural Language Processing	Tom Young,Devamanyu Hazarika,Soujanya Poria, Erik Cambria	Various Deep learning algorithms	E. Cambria and B. White, “Jumping NLP curves: A review of natural language processing research,” IEEE Comput. Intell. Mag., vol. 9, no. 2, pp. 48–57, May 2014.
4	Einstein's Framework for NLP	Ahmida Bendjoudi	Physics for Natural Language Processing	Abhay Ashtekar and Jorge Pullin. The overview chapter in loop quantum gravity: The first 30 years. arXiv preprint arXiv:1703.07396, 2017
5	Inducing brian relevant bias in NLP	Dan Schwatz,Mariya Toneva,Leila Wenbe	BERT	Cichy, R. M., Pantazis, D., and Oliva, A. (2016). Similarity-based fusion of meg and fmri reveals spatio-temporal dynamics in human cortex during visual object recognition. Cerebral Cortex, 26(8), 3563–3579.
6	Arabic Natural Language Processing	Imane Guellil a,Houda Saâdane,Faical Azouaou,Billel Gueni, Damien Nouvel	unigrams, bigrams and trigrams, by using TF-IDF and with SVM classifiers.	I. Guellil, H. Saâdane, F. Azouaou et al., Arabic natural language processing: An overview, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.02.006
7	Fighting post-truth using natural language processing	EstelaSaquete, DavidTomás, PalomaMoreda, PatricioMartínez-Barco, ManuelPalomar	NLP approaches (morphological, syntactic, lexical or semantic parsing) and PRISMA	Bastos, M. T., & Mercea, D. (2019). The brexit botnet and user-generated hyperpartisan news. Social Science Computer Review, 37(1), 38–54. doi:10.1177/ 0894439317734157

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
8	Natural Language Processing Tools to Identify&Classify Periprosthetic Femur Fractures	Meagan E.Tibbo[MD], Cody.C.Wyles[MD],Sunyang Fu[MHI],Sunghwan Sohn[PhD],David G.Lewallen	PPFFx detection, and Vancouver classification	http://refhub.elsevier.com/S0883-5403(19)30709-0/sref1
9	Impact of luxury brand's social media marketing on customer engagement, using NLP	Xia Liua,∗ , Hyunju Shinb , Alvin C. Burnsc	Semantic Analysis	Xia Liu, Hyunju Shin and Alvin C. Burns, Journal of Business Research, https://doi.org/10.1016/j.jbusres.2019.04.042
10	ScispCY: Fast&Robust Models for Biomedical Natural Lnaguage Processing		POS tagging, Named Entity Recognition, Sentence Segmentation and Citation handling	Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew E. Peters, Joanna Power, Sam Skjonsberg, Lucy Lu Wang, Chris Wilhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni. 2018. Construction of the literature graph in semnatic scholar. in NAACL-HLT
11	Review of Natrual Language Processing techniques for opinion mining systems	Shiliang Sun, Chen Luo, Junyu Chen	POS tagging, Named Entity Recognition, Sentence Segmentation and Citation handling	Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information Fusion, 36, 10–25. doi:10.1016/j.inffus.2016.10.004
12	Natural Language Processing of Social Media as Screening for Suicide Risk	Glen Coppersmith, Ryan Leary Patrick Crutchley, Alex Fine	LSTM[Long-Short Time Memory]	Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. Biomedical Informatics Insights, 10, 117822261879286. doi:10.1177/1178222618792860
13	Introduction to Arabic Natural Language Processing	Nizzar.Y.Habash	Hidden Markov Models	Habash, N. Y. (2010). Introduction to Arabic Natural Language Processing. Synthesis Lectures on Human Language Technologies, 3(1), 1–187. doi:10.2200/s00277ed1v01y201008hlt010
14	Comparitive Study of CNN and RNN for Natural Language Processing	Wenpeng Yin, Katharina Kann, Mo Yu, Hinrich Schutze	CNN[Convolution Neural Network] GRU[Gated Recurrent Unit] and LSTM	Heike Adel and Hinrich Schutze. 2017. Exploring dif- ¨ ferent dimensions of attention for uncertainty detection. In Proceedings of EACL.

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
15	Dialouge Games: Metacommunication Structures for Natural Language Processing	James A.Levin, James A.Moore	Dialouge Game Model[DGM]	LEVIN, J., & MOORE, J. (1977). Dialogue-Games: Metacommunication structures for natural language interaction. Cognitive Science, 1(4), 395–420. doi:10.1016/s0364-0213(77)80016-5
16	Learning Language Games through Interaction	Sida I.Wang, Percy Liang, Christopher D.Manning	Semantic Parsing Model	J. Berant, A. Chou, R. Frostig, and P. Liang. 2013. Semantic parsing on Freebase from questionanswer pairs. In Empirical Methods in Natural Language Processing (EMNLP).
17	Playing Text-Adventure Games with Graph-Based Deep Reinforcement Learning	Prithviraj Ammamnabrolu, Mark).Reil	Knowledge Graph DQN(KG-DQN)	Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. 2017. Reading Wikipedia to answer opendomain questions. In Association for Computational Linguistics (ACL).
18	Natural language processing for improving hearing-impaired students reading skills	Lic.Claudia Beatriz Quiroz Pelayo, Dra.Silvia Berenice Fajadro Flores, Dr.Jorge Rafael Gutierrez Pulido	Semantic computation and analysis	Pelayo, C. B. Q., Flores, S. B. F., & Pulido, J. R. G. (2017). Natural Language Processing for Improving Hearing Impaired Student Reading Skills. 2017 International Conference on Information Systems and Computer Science (INCISCOS). doi:10.1109/inciscos.2017.54
19	Using NLP to Automatically Detect Self-Admitted Technical Debt	Everton da S.Maldonado, Emad Shihab, Nikolaos Tsantalis	Latent Semantic Indexing[LSI] and Stanford Classifiers	Maldonado, E. da S., Shihab, E., & Tsantalis, N. (2017). Using Natural Language Processing to Automatically Detect Self-Admitted Technical Debt. IEEE Transactions on Software Engineering, 43(11), 1044–1062. doi:10.1109/tse.2017.2654244
20	Improving Prediction of Suicide and Accidental Death after discharge from Genreal Hospitals	Thomas H.McCoy, Victor M.Castro, Ashlee M.Roberson, Leslie A.Snapper, Roy H.Perils	Charlson Comorbidity Index, Cox Regression	. Niculescu AB, Levey D, Le-Niculescu H, Niculescu E, Kurian SM, Salomon D. Psychiatric blood biomarkers: avoiding jumping to premature negative or positive conclusions. Mol Psychiatry. 2015;20(3):286-288
21	Detecting Phishing Attacks Using Natural Language Processing and Machine Learning	Tianrui Peng,Ian G. Harris,Yuki Sawa	Semantic Analysis	AGGARWAL, S., KUMAR, V., AND SUDARSAN, S. D. Identification and detection of phishing emails using natural language processing techniques. In Proceedings of the 7th International Conference on Security of Information and Networks (2014), SIN '14.

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
22	Inducing brain-relevant bias in natural language processing models	Dan Schwartz,Mariya Toneva,Leila Wehbe	MEG and fMRI data,BERT architecture	Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289–300
23	Natural Language Processing in Oncology	Wen-wai Yim, BS; Meliha Yetisgen, PhD; William P. Harris, MD; Sharon W. Kwan, MD	heuristic algorithms	Yim W, Yetisgen M, Harris WP, Kwan SW. Natural Language Processing in Oncology: A Review. JAMA Oncol. 2016;2(6):797–804. doi:10.1001/jamaoncol.2016.0213
24	Natural Language Processing of Reddit Data to Evaluate Dermatology Patient Experiences and Therapeutics	Edidiong Okon, BSE, Vishnutheja Rachakonda, BS, Hyo Jung Hong, AB, Chris Callison-Burch, PhD, Jules Lipoff, MD	software pipeline preprocessed Reddit comments from 2005 to 2017,d Latent Dirichlet allocation (LDA)	Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011;18(5):544-551.
25	Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment	Daniel J. Feller, Jason Zucker, MD, Michael T Yin, MD, Peter Gordon, MD, and Noémie Elhadad, PhD	EHR processing,Variable Selection and Statistical Modeling	Healthix | Public health information exchange (HIE). Healthix Available at: 433 http://healthix.org/. (Accessed: 2nd August 2016)
26	Deep Bayesian Natural Language Processing	Jen-Tzung Chien	Bayesian Learning,Deep Learning	Emre Aksan and Otmar Hilliges. 2019. STCN: stochastic temporal convolutional networks. In Proc. of International Conference on Learning Representations, page 2019
27	The Stanford CoreNLP Natural Language Processing Toolkit	Christopher D. Manning,Mihai Surdeanu,John Bauer,Jenny Finkel,Steven J. Bethard,David McClosky	Tokenize,POS,Lemma,Regex.NER,Parse,Sentiment	Steven Bethard, Philip Ogren, and Lee Becker. 2014. ClearTK 2.0: Design patterns for machine learning in UIMA. In LREC 2014.
28	Detection of Duplicate Defect Reports Using Natural Language Processing	Per Runeson, Magnus Alexandersson and Oskar Nyholm	Tokenization, Stemming and Stop Words	P. Runeson, M. Alexandersson and O. Nyholm, "Detection of Duplicate Defect Reports Using Natural Language Processing," 29th International Conference on Software Engineering (ICSE'07), Minneapolis, MN, 2007, pp. 499-510, doi: 10.1109/ICSE.2007.32.

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
29	Machine Learning Approach to Fake News Detection Using Knowledge Verification and Natural Language Processing	Marina Danchovsky IbrishimovaKin Fun Li	Recognizing Textual Entailment (RTE),	Ibrishimova, M. D., & Li, K. F. (2019). A Machine Learning Approach to Fake News Detection Using Knowledge Verification and Natural Language Processing. Advances in Intelligent Systems and Computing, 223–234. doi:10.1007/978-3-030-29035-1_22
30	Clinical Natural Language Processing in languages other than English: opportunities and challenges	Aurélie Névéol , Hercules Dalianis, Sumithra Velupillai, Guergana Savova and Pierre Zweigenbaum	Morphology,POS tagging, Parsing, Segmentation	Névéol, A., Dalianis, H., Velupillai, S. et al. Clinical Natural Language Processing in languages other than English: opportunities and challenges. J Biomed Semant 9, 12 (2018). https://doi.org/10.1186/s13326-018-0179-8
31	Jumping NLP Curves: A Review of Natural Language Processing Research	Erik Cambria; Bebo White	Production Rule, Semantic pattern, First Order Logic(FOL), Bayseian&Semantic network, Onotolgy Web Language(OWL)	E. Cambria and B. White, "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]," in IEEE Computational Intelligence Magazine, vol. 9, no. 2, pp. 48-57, May 2014, doi: 10.1109/MCI.2014.2307227.
32	Using natural-language processing to produce weather forecasts	E. Goldberg; N. Driedger; R.I. Kittredge	FOG(A bilingual bilingual report generator that produces routine and special purpose forecast directly from the FPA's graphical weather predictions)	E. Goldberg, N. Driedger and R. I. Kittredge, "Using natural-language processing to produce weather forecasts," in IEEE Expert, vol. 9, no. 2, pp. 45-53, April 1994, doi: 10.1109/64.294135.
33	Natural Language Processing: A Paninian Perspective	Akshar Bharati, Vineet Chaitanya, Rajeev Sangal	Paninian Parser,Tree Adjoining Grammar, Lexical Functional Grammar	Bharati, Akshar & Chaitanya, Vineet & Sangal, Rajeev & Gillon, Brendan. (2002). Natural Language Processing: A Paninian Perspective.
34	Natural Language Processing in Radiology: A Systematic Review	Ewoud Pons , Loes M. M. Braun, M. G. Myriam Hunink, Jan A. Kors	BROK,cTAKES,LEXIMER,MALLET,RadLex®, SAPHIRE,YTEX	Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012;13(6):395–405.
35	A bibliometric analysis of natural language processing in medical research	Xieling Chen, Haoran Xie, Fu Lee Wang, Ziqing Liu, Juan Xu, Tianyong Hao	Speech information recognition , semantic labeling , syntactic parsing , word sense disambiguation , negation detection , and temporal analysis	Chen, X., Xie, H., Wang, F. et al. A bibliometric analysis of natural language processing in medical research. BMC Med Inform Decis Mak 18, 14 (2018). https://doi.org/10.1186/s12911-018-0594-x

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
36	Natural language processing of clinical notes for identification of critical limb ischemia	Naveed Afzal, Vishnu Priya Mallipeddi, Sunghwan Sohn, Hongfan Liu, Rajeev Chaudry, Christopher G.Scott, Iftikhar J.Kullo, Adelaide M Arruda-Olson	Rochester Epidemiology Project (REP) to assemble a community-based PAD cohort from Olmsted County	Afzal, N., Mallipeddi, V. P., Sohn, S., Liu, H., Chaudhry, R., Scott, C. G., … Arruda-Olson, A. M. (2018). Natural language processing of clinical notes for identification of critical limb ischemia. International Journal of Medical Informatics, 111, 83–89. doi:10.1016/j.ijmedinf.2017.12.024
37	Lorentz Transformations & Special Relativity for Natural Language Processing	Ahmida Bendjoudi	Einstein's framework for Natural Language Processing	https://www.researchgate.net/profile/Ahmida_Bendjoudi2/publication/338606833_Lorentz_Transformations_Special_Relativity_for_Natural_Language_Processing/links/5e1f173da6fdcca1f12c6013/Lorentz-Transformations-Special-Relativity-for-Natural-Language-Processing.pdf
38	DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records	Guergana K.Savova, Eugene Tseytlin, Sean Finan, Melissa Castine, TImothy Miller, Olga Medvedeva,David Haris, Harry Hochheiser,Chen Li, Girish Chavan, Rebecca S.Jacobson	DeepPhe system, Apache cTAKES natural language processing system	Savova, G. K., Tseytlin, E., Finan, S., Castine, M., Miller, T., Medvedeva, O., … Jacobson, R. S. (2017). DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records. Cancer Research, 77(21), e115–e118. doi:10.1158/0008-5472.can-17-0615
39	Unlocking echocardiogram measurements for heart disease research through natural language processing	Olga V. Patterson , Matthew S. Freiberg, Melissa Skanderson, Samah J. Fodeh , Cynthia A. Brandt, and Scott L. DuVall	Curated semantic bootstrapping, Veterans Aging CohortStudy (VACS),Corporate Data Warehouse (CDW)	Patterson, O.V., Freiberg, M.S., Skanderson, M. et al. Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC Cardiovasc Disord 17, 151 (2017). https://doi.org/10.1186/s12872-017-0580-8
40	Data science in light of natural language processing: An overview	Imad Zerou, Abdelhak Lakhouja	Hidden Markow Models (HMMs)	Zeroual, I., & Lakhouaja, A. (2018). Data science in light of natural language processing: An overview. Procedia Computer Science, 127, 82–91. doi:10.1016/j.procs.2018.01.101
41	Arabic Natural Language Processing and Machine Learning-Based Systems	Souad Larabi,lmarie-Sainte,Nada Al AlyanIi,Sihaam Aloyaibi ,Sanaa Ghouzali , (Member, IEEE), AND Ibrahim Abunadi , (Member, IEEE)	Naive Bayese, Support Vector Machienes	Larabi Marie-Sainte, S., Alalyani, N., Alotaibi, S., Ghouzali, S., & Abunadi, I. (2019). Arabic Natural Language Processing and Machine Learning-Based Systems. IEEE Access, 7, 7011–7020. doi:10.1109/access.2018.2890076
42	Natural language processing for automated detection of incidental durotomy	Aditya V. Karhade, Michiel E.R. Bongers, MDaOlivier Q. Groot, MDa, Erick R. Kazarian, MDa,b,Thomas D. Cha, MD, MBAa, Harold A. Fogel, MDa,Stuart H. Hershman, MDa, Daniel G. Tobert, MDa,Andrew J. Schoenfeld, MS, MScb, Christopher M. Bono, MDa,James D. Kang, MDb, Mitchel B. Harris, MDa, Joseph H. Schwab, MD, M	extreme gradient-boosting NLP algorithm to detect incidental durotomy,receiver-operating curve (AUC-ROC),current procedural terminology (CPT) and international classification of diseases (ICD)	Karhade, A. V., Bongers, M. E. R., Groot, O. Q., Kazarian, E. R., Cha, T. D., Fogel, H. A., … Schwab, J. H. (2019). Natural language processing for automated detection of incidental durotomy. The Spine Journal. doi:10.1016/j.spinee.2019.12.006

#	Title of the Literature	Author	Methodology/Methods used	References (Name/Link)
43	Improving Patient Cohort Identification Using Natural Language Processing	Raymond Francis Sarmiento,Franck Dernoncourt	Oxford Acute Severity of Illness Score (OASIS),MIMIC is a Medical Information Mart for Intensive Care,	MIT Critical Data. (2016). Secondary Analysis of Electronic Health Records. doi:10.1007/978-3-319-43742-2
44	Use of Natural Language Processing Tools to Identify and Classify Periprosthetic Femur Fractures	Meagan E. Tibbo, MD a, Cody C. Wyles, MD a, Sunyang Fu, MHI b,Sunghwan Sohn, PhD b, David G. Lewallen, MD a, Daniel J. Berry, MD a,Hilal Maradit Kremers, MS, MD	Vancouver Classification,. MedTaggerIE	Tibbo, M. E., Wyles, C. C., Fu, S., Sohn, S., Lewallen, D. G., Berry, D. J., & Maradit-Kremers, H. (2019). Use of Natural Language Processing Tools to Identify and Classify Periprosthetic Femur Fractures. The Journal of Arthroplasty. doi:10.1016/j.arth.2019.07.025
45	Prediction of emergency department patient disposition based on natural language processing of triage notes	Nicholas W. Sterlinga , Rachel E.Patzerb, Mengyu Did , Justin D. Schragera	Bag-of-words (BOW),. Principal components analysis (PCA)	Sterling, N. W., Patzer, R. E., Di, M., & Schrager, J. D. (2019). Prediction of Emergency Department Patient Disposition Based on Natural Language Processing of Triage Notes. International Journal of Medical Informatics. doi:10.1016/j.ijmedinf.2019.06.008
46	Natural language processing of German clinical colorectal cancer notes for guideline-based treatment evaluation	Matthias Beckera,∗ , Stefan Kasperc , Britta Böckmanna,b , Karl-Heinz Jöckelb , Isabel Virchowc	NegEx,Union for International CancerControl (UICC)	Becker, M., Kasper, S., Böckmann, B., Jöckel, K.-H., & Virchow, I. (2019). Natural Language Processing of German Clinical Colorectal Cancer Notes for Guideline-Based Treatment Evaluation. International Journal of Medical Informatics. doi:10.1016/j.ijmedinf.2019.04.022
47	NOBLE – Flexible concept recognition for large-scale biomedical natural language processing	Eugene Tseytlin, Kevin Mitchell, Elizabeth Legowski, Julia Corrigan, Girish Chavan and Rebecca S. Jacobson*	NOBLE Coder,r, cTAKES Dictionary Lookup Annotator(DLA), cTAKES Fast Dictionary Lookup Annotator(FDLA)	Tseytlin, E., Mitchell, K., Legowski, E., Corrigan, J., Chavan, G., & Jacobson, R. S. (2016). NOBLE – Flexible concept recognition for large-scale biomedical natural language processing. BMC Bioinformatics, 17(1). doi:10.1186/s12859-015-0871-y
48	Genetic programming for natural language processing	Lourdes Araujo	Named entity recognition, Hermes Information Extraction Language (HIEL),Relationships identifcation natural language generation (NLG)	Araujo, L. (2019). Genetic programming for natural language processing. Genetic Programming and Evolvable Machines. doi:10.1007/s10710-019-09361-5
49	Feasibility of Natural Language Processing–Assisted Auditing of Critical Findings in Chest Radiology	Marta E. Heilbrun, MDa,1 , Brian E. Chapman, PhDb,1 , Evan Narasimhan, MDc , Neel Patel, MDc, Danielle Mowery, PhD	pyConTextNLP,, Naive Bayes and Support Vector Machines	Heilbrun, M. E., Chapman, B. E., Narasimhan, E., Patel, N., & Mowery, D. (2019). Feasibility of Natural Language Processing–Assisted Auditing of Critical Findings in Chest Radiology. Journal of the American College of Radiology. doi:10.1016/j.jacr.2019.05.038
50	Incorporating natural language processing toimprove classification of axial spondyloarthritis using electronic health records	Sizheng Steven Zhao 1,2,3, Chuan Hong4, Tianrun Cai3,4, Chang Xu3,Jie Huang3, Joerg Ermann3,4, Nicola J. Goodson1,2, Daniel H. Solomon3,4,5,Tianxi Cai4,6, Katherine P. Liao3,4,*	LASSO penalized logistic regression, axSpA,MAP	Zhao, S. S., Hong, C., Cai, T., Xu, C., Huang, J., Ermann, J., … Liao, K. P. (2019). Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. Rheumatology. doi:10.1093/rheumatology/kez375