Heriot Watt University, Dubai

F21AA – Applied Text Analytics

Research Report

Group Number: Dubai Group 3

Student Names and IDs

Sattar Shaikh Eherar (H00450426)
Onafeso Fiyifoluwa (H00466745)
Hariharakumar Rathinar (H00463082)
Yadubanshi Pratibha (H00456474)

Document Control

Item	Description			
Document Title:	**Research Report**			
Doc Ref.	**F21AA – Applied Text Analytics**		Version:	1.0
Classification	○ Public	○ Internal	□ Confidential	○ Confidential & Restricted
Status:	Current	**Type:**	DOC	
Release Date:	20/02/2025			
Revision Date:				

Version No.	Date	Author(s)	Remarks
1.0	21/02/2025	All Team	First Version
2.0	28/02/2025	All Team	Review Comments

Document Review and Approval History

Version No.	Date	Approver(s)	Remarks
2.0	28/02/2025	All Team	Approved

1 Natural Language Processing

Natural Language Processing has evolved as a branch of Artificial Intelligence, devoted to making machines understand human written and spoken language. With the advent of the latest technologies, it has further segregated to 2 fields viz. Natural Language Understanding and Natural Language Generation [2]. As the NLP advances, we attempt to articulate the history of its progression through 4 different eras.

First-era, between 1950 and 1969, started with the intention to obtain word level machine translations using lookups. Most notably, the 1954 George Town University-IBM experiment of converting Russian Language to English [6], showcased potential of language processing by computers and peaked interest of public.

The second era, between 1970 and 1992, rolled out rule-based systems showcasing sophistication and depth in handling complex human language. Primary examples of this are SHRDLU by Terry Winograd and LUNAR by Bill Woods [4]. At this time, these demonstrations were considered significant achievements in the field of Linguistics and knowledge-based AI. They demonstrated a procedural style of a coherent computational system, though lacking robustness and scalability. Towards the late 80's, using grammatic-logical approaches and parsers more commercial systems were developed (e.g. Alvey Natural Language tools [5]).

However, during the third era, between 1993 and 2012, with increased availability of digital text and computational capabilities, NLP transformed from rule-based methods to statistical and machine learning approaches [1]. This led to the development of annotated linguistic resources and named entity datasets, enabling supervised learning and information extraction using Support Vector Machines, Hidden Markov Models etc. At the end of millennium, RNN-Recurrent Neural Networks with LSTM-Long Short-Term Memory networks using word embeddings achieved enhanced machine translation and sentiment analysis.

The fourth era, from 2013 until today, saw the introduction of deep learning and artificial neural network methods. They achieved generalization and improved performance by leveraging vector space rather than symbolic representations representing words and sentences in high-dimensional vector spaces. Transformative shift occurred in 2018 with launch of large-scale self-supervised neural networks like BERT and GPT. These transformer-based models use self-attention mechanisms to process input data in parallel, making them ideal candidates for effective NLP processing [3]. This self-supervised approach revolutionized NLP by enabling parallel processing and state-of-the-art performance. Multimodal AI systems using image, speech and text (ChatGPT) and Reinforced Learning models (DeepSeek) have reignited AI generalization particularly for NLP.

NLP started with rule-based systems, move to statistical methods, and today relies on advanced deep learning models. Focus is now on making NLP systems user-friendly, understandable, bias and hallucination free and its implementations in real world applications.

2 References

1. **P. Manning,** “The new networks of knowledge,” *Dædalus*, vol. 151, no. 2, pp. 125–139, Spring 2022. [Online]. Available: https://tinyurl.com/ydnkdf7z
2. **J. Schmidhuber,** "Deep learning in neural networks: An overview," *arXiv preprint arXiv:1708.05148*, 2017. [Online]. Available: https://arxiv.org/pdf/1708.05148
3. **S. Ruder,** "A Review of the Neural History of Natural Language Processing," *ruder.io*, Oct. 1, 2018. [Online]. Available: https://tinyurl.com/5n7n9jbj
4. **K. S. Jones,** "A history of stretching and compressing words," *University of Cambridge, Computer Laboratory*, 2004. [Online]. Available: https://tinyurl.com/4jz5ckax
5. **Natural Language and Information Processing Group,** "A New Look at ANLT," *University of Cambridge, Computer Laboratory*. [Online]. Available: https://tinyurl.com/2a5yvtkt
6. **W. J. Hutchins,** "The Georgetown-IBM Experiment Demonstrated in January 1954," in *Machine Translation: From Real Users to Research*, Washington, DC, USA, Sep. 28–Oct. 2, 2004, Lecture Notes in Computer Science, vol. 3265, H. L. Somers, Ed. Berlin, Germany: Springer, 2004, pp. 102–114. [Online]. Available: https://tinyurl.com/bdzyh9uw