Heriot Watt University, Dubai

F21AA – Applied Text Analytics

Research Report – Group 3

Document Control

Item	Description			
Document Title:	**Submission 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

SHARED – CONFIDENTIAL / خاص – مشترك

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. 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, 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 were SHRDLU by Terry Winograd and LUNAR by Bill Woods. 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).

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. This led to the development of annotated linguistic resources and named entity datasets, enabling supervised learning information extraction. Hidden Markov Models (HMMs), Support Vector Machines (SVMs), and Conditional Random Fields (CRFs) played key roles in improving these tasks. In the late 2000s, using word embeddings (for better semantic representation), Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) networks were able to achieve 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. This self-supervised approach revolutionized NLP by enabling parallel processing and state-of-the-art performance across tasks. Recent advancements include instruction-tuned models like ChatGPT, multimodal AI systems integrating text, image, and speech, and Reinforcement Learning from Human Feedback (RLHF), which refines model responses for better coherence and alignment with human intent.

NLP has come a long way, starting with rule-based systems, moving to statistical methods, and now relying on advanced deep learning models like Transformers. This shift highlights how quickly the field has advanced. Looking ahead, researchers are focusing on making NLP systems easier to understand, reducing biases, and ensuring they work well in real-world situations.

SHARED – CONFIDENTIAL / مشترك – خاص

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/24bauped
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: