

Kent Academic Repository

Full text document (pdf)

Citation for published version

(2022) Natural Language Engineering: Methods, Tasks and Applications. [Edited Journal]

DOI

<https://doi.org/10.3390/fi14040106>

Link to record in KAR

<https://kar.kent.ac.uk/93867/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>



Editorial

Special Issue “Natural Language Engineering: Methods, Tasks and Applications”

Massimo Esposito ¹, Giovanni Luca Masala ², Aniello Minutolo ¹ and Marco Pota ^{1,*}

¹ Institute for High Performance Computing and Networking–National Research Council of Italy (ICAR-CNR), 80100 Naples, Italy; massimo.esposito@icar.cnr.it (M.E.); aniello.minutolo@icar.cnr.it (A.M.)

² School of Computing, University of Kent, Canterbury CT2 7NZ, UK; g.masala@kent.ac.uk

* Correspondence: marco.pota@icar.cnr.it

Natural language engineering includes a continuously enlarging variety of methods for solving natural language processing (NLP) tasks within a pervasive number of applications. In this field, impressive achievements have been reached recently, by means of systems using deep learning or different approaches, which allowed AI to advance toward human levels in NLP tasks such as translation [1], reading comprehension [2,3], information retrieval [4], and sentiment analysis [5–7], and to build systems for question answering [8–11], conversational systems [12,13], and recommender systems (<https://developers.google.com/machine-learning/recommendation>, accessed on 24 March 2022).

However, despite the remarkable successes in different NLP tasks, natural language engineering is nowadays a field of research of increasing interest due to the remaining difficulties associated with its comprehension and generation, which are capabilities of humans still not well understood by computer systems from a cognitive perspective. Current difficulties include the complexity of deep learning models, growing in directions chosen empirically [14], the difficulty of scaling them down for implementation on the edge, the scarcity of datasets for some languages (<https://www.aclweb.org/portal/content/emnlp-workshop-deep-learning-low-resource-nlp>, accessed on 24 March 2022), and the lack of explainability of the models [15].

This Special Issue highlights the most recent research being carried out in the field of NLP methods, to face these open issues, with particular emphasis on emerging approaches for learning interactively or autonomously from data, single and multiple language understanding and grounding for analysis and generation, as well as potential or real applications in different domains and everyday devices.

To this aim, this Special Issue gathers original contributions by researchers with broad expertise in various fields—natural language processing, cognitive science and psychology, artificial intelligence and neural networks, computational modeling and neuroscience—discussing their cutting-edge work as well as perspectives on future directions in the whole range of theoretical and practical aspects, technologies, and systems in this research area.

There are six contributions selected for this Special Issue, representing progress and potential applications in the following NLP areas specifically addressed:

1. **Low-resource natural language processing.** Yimam et al. state that the available pre-trained models do not fit well with the need for low-resource languages; thus, they introduce different semantic models for Amharic and fine-tune two pre-trained models and train seven new models. Moreover, they employ these models for different NLP tasks and study their impact.
2. **Natural language understanding, generation and grounding.** Agafonova et al. revisit the receptive theory in the context of computational creativity; they present a fully autonomous text generation engine with raw output simulating the narrative of



Citation: Esposito, M.; Masala, G.L.; Minutolo, A.; Pota, M. Special Issue “Natural Language Engineering: Methods, Tasks and Applications”. *Future Internet* **2022**, *14*, 106. <https://doi.org/10.3390/fi14040106>

Received: 2 March 2022

Accepted: 23 March 2022

Published: 26 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

a mad digital person and discuss the impact of receptive theory, chance discovery, and simulation of fringe mental state on the understanding of computational creativity.

3. **Neuroscience-inspired cognitive architectures.** Onorati et al. propose a model to control a specific class of syntax-oriented neural networks by adding declarative rules, by exploiting parse trees and subtrees, to include human control in NLP systems, and they show that declarative rules representing human knowledge can be effective for some NLP tasks.
4. **Search and information retrieval.** Yu et al. underline that classification of resource can help the filtering of massive resources, and they propose for this scope an Association Content Graph Attention Network, which is based on association features and content attributes of academic resources, considering both semantic relevance and academic relevance, to improve the accuracy of academic resource classification.
5. **Text de-identification.** Libbi et al. consider the lack of large, annotated Electronic Health Records datasets due to privacy concerns and annotation costs, thus they propose the use of language models for generating artificial data jointly with annotations that can be effectively used, alone or in combination with real data, to train supervised named-entity recognition models for de-identification.
6. **Applications in science, engineering, medicine, healthcare, finance, business, law, education, industry, transportation, retailing, telecommunication and multimedia.** Song and Huang propose to use the massive amount of data generated by social media for disaster analysis, and in particular to use Twitter to track disaster events to make a speedy rescue plan, and for this scope, they propose a sentiment-aware contextual model, consisting of a layer that can generate sentimental contextual embeddings from tweets, a BiLSTM layer with attention, and a 1D convolutional layer for local feature extraction, demonstrating superior performance in Tweets-based disaster analysis.

Acknowledgments: This Special Issue was successful thanks to the valuable contributions of all the authors, the dedicated referees, and the Editorial team of *Future Internet*.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tan, Z.; Wang, S.; Yang, Z.; Chen, G.; Huang, X.; Sun, M.; Liu, Y. Neural machine translation: A review of methods, resources, and tools. *AI Open* **2020**, *1*, 5–21. [[CrossRef](#)]
2. Guarasci, R.; Minutolo, A.; Damiano, E.; De Pietro, G.; Fujita, H.; Esposito, M. ELECTRA for Neural Coreference Resolution in Italian. *IEEE Access* **2021**, *9*, 115643–115654. [[CrossRef](#)]
3. Marulli, F.; Pota, M.; Esposito, M. A comparison of character and word embeddings in bidirectional LSTMs for POS tagging in Italian. In *Intelligent Interactive Multimedia Systems and Services*; De Pietro, G., Gallo, L., Howlett, R.J., Jain, L.C., Vlacic, L., Eds.; Springer: Berlin, Germany, 2019; Volume 98, pp. 14–23.
4. Guarasci, R.; Damiano, E.; Minutolo, A.; Esposito, M.; De Pietro, G. Lexicon-grammar based open information extraction from natural language sentences in Italian. *Expert Syst. Appl.* **2020**, *143*, 112954. [[CrossRef](#)]
5. Pota, M.; Ventura, M.; Fujita, H.; Esposito, M. Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets. *Expert Syst. Appl.* **2021**, *181*, 115119. [[CrossRef](#)]
6. Pota, M.; Esposito, M.; Palomino, M.A.; Masala, G.L. A subword-based deep learning approach for sentiment analysis of political tweets. In Proceedings of the 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), Krakow, Poland, 16–18 May 2018.
7. Yadav, A.; Vishwakarma, D.K. Sentiment analysis using deep learning architectures: A review. *Artif. Intell. Rev.* **2020**, *53*, 4335–4385. [[CrossRef](#)]
8. Pota, M.; Fuggi, A.; Esposito, M.; De Pietro, G. Extracting Compact Sets of Features for Question Classification in Cognitive Systems: A Comparative Study. In Proceedings of the 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, Poland, 4–6 November 2015.
9. Esposito, M.; Damiano, E.; Minutolo, A.; De Pietro, G.; Fujita, H. Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering. *Inf. Sci.* **2020**, *514*, 88–105. [[CrossRef](#)]
10. Pota, M.; Esposito, M.; De Pietro, G.; Fujita, H. Best Practices of Convolutional Neural Networks for Question Classification. *Appl. Sci.* **2020**, *10*, 4710. [[CrossRef](#)]
11. Yuan, S.; Zhang, Y.; Tang, J.; Hall, W.; Cabotà, J.B. Expert finding in community question answering: A review. *Artif. Intell. Rev.* **2020**, *53*, 843–874. [[CrossRef](#)]

12. Minutolo, A.; Esposito, M.; De Pietro, G. A conversational chatbot based on knowledge-graphs for factoid medical questions. In Proceedings of the 16th International Conference on Intelligent Software Methodologies, Tools and Techniques, KitaKyushu, Japan, 26–28 September 2017.
13. Minutolo, A.; Damiano, E.; De Pietro, G.; Fujita, H.; Esposito, M. A conversational agent for querying Italian Patient Information Leaflets and improving health literacy. *Comput. Biol. Med.* **2021**, *141*, 105004. [[CrossRef](#)]
14. Pota, M.; Marulli, F.; Esposito, M.; De Pietro, G.; Fujita, H. Multilingual POS tagging by a composite deep architecture based on character-level features and on-the-fly enriched Word Embeddings. *Knowl. Based Syst.* **2019**, *164*, 309–323. [[CrossRef](#)]
15. Zohuri, B.; Moghaddam, M. Deep Learning Limitations and Flaws. *Mod. Approaches Mater. Sci. Short Commun.* **2020**, *2*, 241–250. [[CrossRef](#)]