Semantics in the Deep: Semantic Analytics for Big Data

Dimitrios Koutsomitropoulos [1,*], Spiridon Likothanassis [1] and Panos Kalnis [2]

[1] Department of Computer Engineering & Informatics, University of Patras, Patras 26504, Greece; likothan@ceid.upatras.gr

[2] Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia; panos.kalnis@kaust.edu.sa

* Correspondence: kotsomit@ceid.upatras.gr

One cannot help but classify the continuous birth and demise of Artificial Intelligence (AI) trends into the everlasting theme of the battle between connectionist and symbolic AI. During the past decade, the Semantic Web has brought a flavor of intelligence into everyday web browsing and has remained a hot research topic for years. The advent of big data has posed challenges in its scalability; at the same time, it has offered fertile ground for machine learning and computational intelligence techniques to flourish and to optimize, in previously intractable fields. But a remnant of the disbelief that meant such methods were shunned in the past, when compared to the sound foundations of logic, still remains, despite their aweing achievements. As a result, two fundamental questions naturally arise:

- How can semantic technologies contribute towards big data analysis?
- What is the relationship between Semantic Web logical formalisms and automated- and deep-learning techniques?

This has been exactly the reasoning behind the Semantic Analytics for Big Data (SEDSEAL) international workshop [1]. The theme of this special issue has been inspired by and announced at SEDSEAL 2018, in Rhodes, Greece. The interest spiked by this convention of researchers and practitioners has motivated us to pursue its capitalization in the form of a journal publication.

Therefore, the purpose of this special issue is to seek and validate answers to the above questions by putting together a selection of high-quality research and development efforts that consider and present contributions towards how Semantic Web technologies can help to implement and enhance big data analytics. This can be achieved either by extracting value out of these data (e.g., through reasoning), creating sustainable ontology models, offering a solid foundation for deploying learning techniques, or anything in between.

One of the most common user experiences involves web search, and a lot of big data can be generated in the process. The role of neural networks and deep learning as a means to strike a balance between user aspiration, sponsored search and big data generators is thoroughly reviewed in "Neural Networks in Big Data and Web Search" [2]. Among others, web ranking, learning to rank and semantic recommendation algorithms and systems are presented, and additional deep learning extensions are being examined.

Web-based knowledge management is considered in the paper "CRC806-KB: A Semantic MediaWiki Based Collaborative Knowledge Base for an Interdisciplinary Research Project" [3]. To tackle interoperability issues between disparate knowledge domains and aid collaborative contributions of field experts, the authors deploy semantic extensions on top of a wiki collaboration platform on the Web. Then, they suggest additional tooling to facilitate automated imports of datasets with variable dimensions.

Natural language text is a frequent and often problematic source of big data. Linguistic analysis of large text corpora is in fact semantic analysis, and this is eloquently addressed in the paper "From a Smoking Gun to Spent Fuel: Principled Subsampling Methods for Building Big Language Data Corpora from Monitor Corpora" [4]. The author presents an approach to validate and generalize a downsampling method for efficiently analyzing large text sources and apply it to data pertaining to sociotechnical issues like energy.

We are grateful to the various authors that trusted their work with this special issue. Many thanks are also due to Jingjing Lu and Cody Peng, managing and assistant editors respectively at MDPI, who have greatly helped us to pull off this work. They could not have done a better job at administering all publishing details and letting us focus on the academic part. We are also obliged to the people that have served as members of the editorial review board (in alphabetical order): Paulo Carrasco, Roberto Cerchione, Liviu-Adrian Cotfas, Stefan Gindl, Sabina Necula, Filipe Portela, John Samuel and Paweł Ziemba. Their eager participation in the review process, the sharing of their time and the contribution of their expertise have been of outmost importance for realizing this issue.

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

References

1. Semantics in the Deep: Semantic Analytics for Big Data. Available online: https://sedseal2018.ceid.upatras.gr (accessed on 25 April 2019).
2. Serrano, W. Neural Networks in Big Data and Web Search. *Data* **2019**, *4*, 7. [CrossRef]
3. Willmes, C.; Viehberg, F.; Esteban Lopez, S.; Bareth, G. CRC806-KB: A Semantic MediaWiki Based Collaborative Knowledge Base for an Interdisciplinary Research Project. *Data* **2018**, *3*, 44. [CrossRef]
4. Tidwell, J.H. From a Smoking Gun to Spent Fuel: Principled Subsampling Methods for Building Big Language Data Corpora from Monitor Corpora. *Data* **2019**, *4*, 48. [CrossRef]