

*Editorial Introduction to the Special Articles in the Spring Issue*

# Semantics for Big Data

*Frank van Harmelen, James A. Hendler, Pascal Hitzler, Krzysztof Janowicz*

■ *This editorial introduction summarizes the seven guest-edited contributions to AI Magazine that explore opportunities and challenges arising from transferring and adapting semantic web technologies to the big data quest.*

One of the key challenges in making use of big data lies in finding ways of dealing with heterogeneity, diversity, and complexity of the data, while its volume and velocity forbid solutions available for smaller data sets as based, for example, on manual curation or manual integration of data. Semantic web technologies (Hitzler, Krötzsch, and Rudolph 2010) are meant to deal with these issues, and indeed since the advent of linked data (Bizer, Heath, and Berners-Lee 2009) a few years ago, they have become central to mainstream semantic web research and development. We can easily understand linked data as being a part of the greater big data landscape, as many of the challenges are the same (Hitzler and Janowicz 2013). The linking component of linked data, however, puts an additional focus on the integration and conflation of data across multiple sources.

At the AAAI Fall Symposium in November 2013, we explored opportunities and challenges arising from transferring and adapting semantic web technologies to the big data



**The Twenty-Eighth  
Annual Conference on  
Innovative Applications of  
Artificial Intelligence  
(IAAI-16)**

**February 12–17, 2016  
Phoenix, Arizona USA**

*Please Join Us!*  
[www.aaai.org/iaai16](http://www.aaai.org/iaai16)

quest. This issue of *AI Magazine* is a follow-up from that meeting and contains significantly extended, enhanced, and updated contributions. We summarize the articles in the following paragraphs.

The introductory article, *Why the Data Train Needs Semantic Rails*, by Krzysztof Janowicz, Frank van Harmelen, James A. Hendler, and Pascal Hitzler, discusses the opportunities and challenges related to combining formal semantics with data analytics approaches. It is followed by *Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation* by Lora Aroyo and Chris Welty, which discusses how to deal with subjectivity using a crowd-based approach.

*Exploiting Semantics for Big Data Integration* by Craig A. Knoblock and Pedro Szekely shows how to use semantics to integrate heterogeneous data in a cultural heritage use case, while *Semantics-Empowered Big Data Processing with Applications* by Krishnaprasad Thirunarayan and Amit Sheth addresses the role of semantics in analyzing and processing big data that arises in the context of physical-cyber-social systems.

*Early Steps Toward Web Scale Information Extraction with LODIE* by Anna Lisa Gentile, Ziqi Zhang, and Fabio Ciravegna describes a linked-data-based methodology for web scale information extraction. *Using Semantics and Statistics to Turn Data into Knowledge* by Jay Pujara, Hui Miao, Lise Getoor, and William W. Cohen describes the use of ontological constraints to eliminate errors when deriving knowledge from text.

The final article, *Entity Type Recognition for Heterogeneous Semantic Graphs*, by Jennifer Sleeman, Tim Finin, and Anupam Joshi addresses coreference resolution using a supervised machine-learning approach.

**References**

Bizer, C.; Heath, T.; and Berners-Lee, T. 2009. Linked Data — The Story So Far. *International Journal on Semantic Web and Information Systems* 5(3): 1–22.

Hitzler, P., and Janowicz, K. 2013. Linked Data, Big Data, and the Fourth Paradigm. *Semantic Web* 4(3): 233–235.

Hitzler, P.; Krötzsch, M.; and Rudolph, S. 2010. *Foundations of Semantic Web Technologies*. Boca Raton, FL: Chapman & Hall / CRC.

**Frank van Harmelen** is a professor in knowledge representation and reasoning at the Vrije Universiteit Amsterdam, The Netherlands.

**James A. Hendler** is a professor for artificial intelligence at the Rensselaer Polytechnic Institute, USA.

**Pascal Hitzler** is an associate professor for semantic web technologies at the Wright State University, USA.

**Krzysztof Janowicz** is an assistant professor for geoinformatics at the University of California, Santa Barbara, USA.