



54<sup>th</sup> CIRP Conference on Manufacturing Systems

# Knowledge Discovery in Heterogeneous and Unstructured Data of Industry 4.0 Systems: Challenges and Approaches

Simon Kamm<sup>a,\*</sup>, Nasser Jazdi<sup>a</sup>, Michael Weyrich<sup>a</sup>

<sup>a</sup>*Institute of Industrial Automation and Software Engineering, University of Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany*

\* Corresponding author. Tel.: +49 711 685 67293; fax: +49 711 685 67302. *E-mail address:* [simon.kamm@ias.uni-stuttgart.de](mailto:simon.kamm@ias.uni-stuttgart.de)

---

## Abstract

With the rise of the Internet of Things and Industry 4.0, the number of digital devices and their produced data increases tremendously. Due to the heterogeneity of devices, the generated data is mostly heterogeneous and unstructured. This challenges established approaches for knowledge discovery, which typically consume structured data from one source. The paper first describes aspects of data heterogeneity and their relevance for Industry 4.0 systems. Following, the upcoming challenges for different steps inside the knowledge discovery process for Industry 4.0 systems, such as for data integration and data mining, are discussed. Additionally, it mentions approaches to tackle them.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the 54<sup>th</sup> CIRP Conference on Manufacturing System

*Keywords:* : Heterogeneous Data, Industry 4.0, Internet of Things, Knowledge Discovery, Machine Learning, Semantic Data Integration, Unstructured Data

---