



Contribution ID: 335

Type: **Talk**

[55] AI assisted Scalable Knowledge Ingestion for Automated Discoveries

Thursday 29 August 2019 15:10 (20 minutes)

Over the past few decades, the amount of scientific articles and technical literature has increased exponentially in size. Consequently, there is a great need for systems that can ingest these documents at scale and make the contained knowledge discoverable. Unfortunately, both the format of these documents (e.g. the PDF format or bitmap images) as well as the presentation of the data (e.g. complex tables and figures) make the extraction of qualitative and quantitative data extremely challenging. In this talk, we will present our three pronged approach to this problem and show practical examples in the field of Material Science and Oil&Gas. We will start by introducing a scalable service [1] that is able to ingest documents at scale and exploits state-of-the-art AI models to obtain very high accuracies. Next, we will show how the data contained in the ingested documents can be extracted using NLP methods. Finally, we will show how the extracted data can be efficiently queried using Knowledge Graphs and how one can obtain new insights from these graphs by applying advance analytics [2].

[1] https://www.researchgate.net/publication/325359423_Corpus_Conversion_Service_A_Machine_Learning_Platform_to_Ingest_Docu

[2] https://www.researchgate.net/publication/303551320_Stochastic_Matrix-Function_Estimators_Scalable_Big-Data_Kernels_with_High_Performance

Author: Dr DOLFI, Michele (IBM Research - Zurich)

Presenter: Dr DOLFI, Michele (IBM Research - Zurich)

Session Classification: Quantum and Artificial Intelligence: New Jobs for Physicists in Emergent Industries