Towards the Bosch Materials Science Knowledge Base

J. Strötgen, T.K. Tran, A. Friedrich, D. Milchevski, F. Tomazic, A. Marusczyk,
H. Adel, D. Stepanova, F. Hildebrand, and E. Kharlamov

Robert Bosch GmbH, Corporate Research and Bosch Center for Artificial Intelligence, Germany

Materials Science Knowledge: Finding Diamonds in the Dirt. For manufacturing companies, employing new and innovative materials is crucial for developing competitive products. There are thousands of materials available for production purposes within the automotive industry, for consumer goods, energy solutions, or building technology, all fields in which the Robert Bosch GmbH is an active player. In order to respond to new demanding requests from the market or regulatory organisations, Bosch engineers constantly introduce new materials that meet complex requirements.

Developing new materials critically depends on the ability to find high quality answers about existing materials in a timely manner. In the last decades, there has been an exponential growth in the volume of information about new materials and chemical components, with thousands of new papers and patents appearing every year. Analyzing this data and finding information relevant to a concrete need is a challenging task for materials engineers and researchers. For example, the following query expresses such an information need: "*Find anode materials in Intermediate Temperature Solid Oxide Fuel Cells (IT-SOFC) that produce high power density.*"

Bosch Knowledge System (BoschKS) for Finding Materials Science Knowledge. In order to support materials science engineers in their information search, we are developing a system (see Fig. 1) fulfilling the following criteria. *(i)* The system integrates information from different sources in a unified Knowledge Graph (KG), i.e., we combine information from relational databases with textual information, relying on the Ontology Data Access technology and existing and novel in-house Natural Language Processing (NLP) techniques, respectively. *(ii)* Besides standard KG search capabilities, it offers complex query answering facilities that support aggregation of information and multi-hop reasoning. *(iii)* It computes provenance for query answers as well as their justifications via the reasoning steps, thus, making answers explainable.

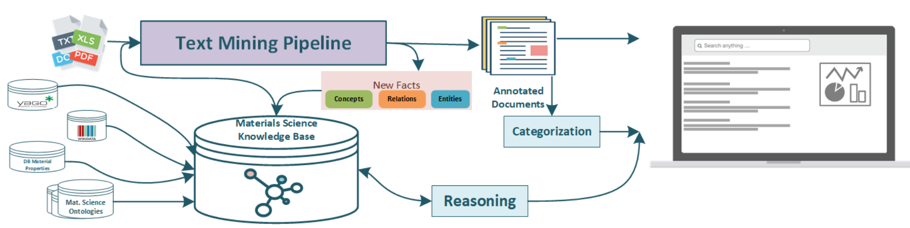

Fig. 1. BoschKS architecture

The Bosch Materials Science Knowledge Base (MatKB) exposes materials science knowledge stored in documents and databases, providing an ontology containing background domain knowledge, and a unified view for information from various data sources. At the moment the ontology contains approximately 8K classes and properties

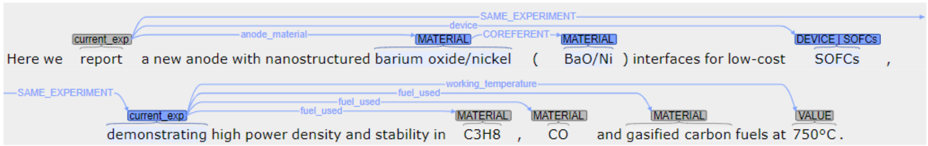

Fig. 2. A sentence from a materials science paper annotated with Experiment frame information.

and 9K axioms. In addition, MatKB's KG consists of 40K facts/triples about materials, growing rapidly. We next briefly discuss some aspects of our system.

From Text to Knowledge. A key component of our system uses NLP techniques for acquiring high quality knowledge from unstructured formats, i.e., for extracting information from scientific publications and patents. We combine different approaches to text mining: *(i)* we make use of purchased in-domain preprocessing components and extract simple cases in a rule-based way; *(ii)* we adapt general-purpose NLP tools to our domain; and *(iii)* we are working with domain experts to create manually annotated high quality reference data; this data is then used to train neural networks to obtain large-scale knowledge. As a first step, we address the use case of extracting fine-grained information about a particular experimental domain from publications. We have developed an annotation scheme for marking up information about scientific experiments which is inspired by frame semantics. We first apply the scheme to experiments on solid oxide fuel cells (SOFCs): we identify words such as "report" (Fig. 2) that indicate that some experiment is being described, and link to the entities that are part of the experiment such as "BaO/Ni" along with the information that this material is used for the SOFC's anode in this case. It is worth noting that this annotation scheme can easily be adapted to other experimental domains, and the collected annotations can be used as seed data in other use cases. In addition, we are working on extracting information that is common to all disciplines within the materials sciences, e.g., microstructural properties, measurement conditions, synthesis procedures or processing.

Knowledge Management and Exploration. The knowledge extracted from the NLP components is stored as triples in MatKB using Stardog. We use the R2RML language to make existing information in relational databases available as virtual graphs. In the front end, a customized version of Metaphactory provides a user-friendly query interface facilitating standard key word search and semantic based faceted search. Furthermore, we have been developing advanced reasoning techniques to improve the quality of MatKB, e.g., by consistency checking and constraint validation, and to guarantee smooth incorporation of the background knowledge ontology and facts in the knowledge graph. For example, the information extracted from the sentence in Fig. 2 is - on its own - insufficient to answer the above query because the type of SOFC being investigated (IT-SOFC) is not mentioned explicitly. However, from background knowledge, we know that the working temperature of IT-SOFCs is usually in the range of $600-800°C$. Such information can be combined with the extracted facts via query rewriting or materialization to provide the answer "BaO/Ni" as a desired material.

Outlook. The development of BoschKS and the Bosch MatKB is ongoing work. This paper describes our first but significant steps towards this goal. We envision a substantial growth of MatKB in the near future integrating data from a broad range of sources. Moreover, we work on integrating more in-house solutions, e.g., rule and ontology learning, into the existing framework. MatKB is not publicly available, but we plan to prepare its public demo version. Finally, we plan to use BoschKS not only for the materials science domain, but further extend it to other domains.