

Context Aware Information Retrieval from Materials Publications

ICME 2016

Sapan Shah*, Dhwani Vora, B P Gautham and Sreedhar Reddy

* Tata Consultancy Services Ltd.,
Pune, India

e-mail: {sapan.hs, vora.dhwani, bp.gautham, sreedhar.reddy}@tcs.com,
web page: <http://www.tcs.com>

ABSTRACT

Knowledge of material properties as a function of material composition and manufacturing process parameters is of significant interest to materials scientists and engineers. A large amount of information of this nature is available in publications especially in the form of experimental measurements, simulation outcomes, etc. In a typical problem solving context, requiring information not available in standard databases, an engineer has to first go through a large collection of publications, either internal reports or open literature, to filter the right set of documents, containing information relevant to the context. Having filtered the publications, an engineer has to then go through each of them to extract the relevant pieces of information. Our goal is to help automate some of these steps. In this paper, we present our ongoing work on a system that provides information search and extraction based on material entities such as elemental composition, manufacturing processes (along with processing parameters) and properties. The repository of materials publications/reports is first pre-processed to extract material entities and their values. The extracted entities as well as the publication content are then indexed for use by the search system. To specify search criteria, the system provides a domain specific query language and also supports “and”, “or” and “not” operators for facilitating complex search queries. The accuracy of our system critically depends on the accuracy of the information extraction component. To address this, we are developing a sophisticated extraction component that combines rule based and machine learning based approaches for extracting different material entities. In addition to providing search, the extracted entities can also be stored in a materials data repository. We have conducted an experiment on a small library of publications on steel on which searches such as “get the list of publications about steel which has carbon composition between 0.2 and 0.3 and on which tempering is performed for about 30 to 40 min” are performed. We compare our system with a keyword based search approach.