Semantic Federation of Product Information from Structured and Unstructured Sources

Matthias Wauer, Technische Universität Dresden, Germany
Johannes Meinecke, SAP Research Dresden, Germany
Daniel Schuster, Technische Universität Dresden, Germany
Andreas Konzag, BMW Group, Germany
Markus Aleksy, ABB Corporate Research, Germany
Till Riedel, Karlsruhe Institute of Technology, Germany

ABSTRACT

Product-related information can be found in various data sources and formats across the product lifecycle. Effectively exploiting this information requires the federation of these sources, the extraction of implicit information, and the efficient access to this comprehensive knowledge base. Existing solutions for product information management (PIM) are usually restricted to structured information, but most of the business-critical information resides in unstructured documents. We present a generic architecture for federating heterogeneous information from various sources, including the Internet of Things, and argue how this process benefits from using semantic representations. A reference implementation tailor-made to business users is explained and evaluated. We also discuss several issues we experienced that we believe to be valuable for researchers and implementers of semantic information systems, as well as the information retrieval community.

Keywords: Federated Information Systems, Information Extraction, Information Systems, Knowledge Management, Ontology, Product Information Management, Semantic Web

INTRODUCTION

Product-related information is generated, accessed and manipulated along the product lifecycle in heterogeneous formats. Only part of this information can be accessed using state-of-the-art product information systems as large parts of this information are only available in unstructured sources or distributed along different databases and legacy systems. The challenge to create an all-embracing view on products is huge. Such a comprehensive product information system has to integrate and harmonize data from all phases of the product lifecycle, all different source formats.

DOI: 10.4018/jbdcn.2011040105
like unstructured documents, sensor information or product databases. Furthermore, it must even cross organization boundaries as different stakeholders may be responsible for the design, production, delivery, and service of a product.

The Aletheia project (Aletheia, 2009) is a unique attempt to bring together industry partners (ABB, BMW, Deutsche Post DHL, Otto, SAP) with five innovative application scenarios from different phases of the product lifecycle and five different landscapes of current state-of-the-art product information management. All these partners have a keen interest in improving the information flow internally as well as with their customers and partners and to open up new sources of product-related information like Web 2.0 pages.

In this paper we try to answer the research question if it is possible to federate structured as well as unstructured sources of product information along the product lifecycle. We use semantic technologies for this purpose and deploy and advance information extraction techniques. The scenarios describe two of the use cases of the Aletheia project clarifying the opportunities of federated product information systems (FPIS). We further discuss requirements derived from these and other scenarios. A discussion of existing architectures for semantic information management and federation shows the need for a new architecture matching the requirements mentioned. The contributions of this paper consist of:

1. A discussion of design decisions for FPIS,
2. A high-level component architecture for FPIS, including a concept for data sharing between organizations,
3. A detailed concept of the Aletheia Service Hub, our central component for information federation within organizations,
4. A reference implementation of a semantic FPIS.

We conclude with a discussion of the results achieved so far.

SCENARIOS AND REQUIREMENTS

In order to motivate our research, we discuss two scenarios in the industrial sector. They are derived from two case studies conducted in the Aletheia project, focusing on:

1. Product lifecycle management (PLM) at ABB, a large company providing power and automation products, technology, and service¹, and
2. Knowledge management in automotive engineering at the BMW company.

Use Case ABB

The customer has installed several products of the company at their local site. A service team of the customer notices that one of the devices is defective. Even though they have the knowledge which devices are applied at this installation, they lack the capability of identifying the actual cause and repairing the device. Hence, they contact the company’s call center that records the service request. However, neither the customer’s service team nor the call center associate has expert knowledge about the defective device. The service report therefore is frequently inaccurate, and preparing the service operation is laborious for an assigned service technician on the basis of this report.

On site, most of the suitable unstructured information is not consulted by the service technician because finding it based on the available information is cumbersome. If additional spare parts are required to repair the device, the service technician has to manually coordinate the order of a spare part and its delivery with the company’s call center, the logistics provider, and the customer’s service team. This causes several phone calls and requires much effort because the service technician’s available information is not integrated with those of the other parties.

This use case can be optimized by three means. First, the customer should be able to
Data Processing and Exchange Challenges in Video-Based Wireless Sensor Networks
