Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 47 (2016) 294 – 299

www.elsevier.com/locate/procedia

Product-Service Systems across Life Cycle

"A Semantic Information Retrieval Framework within the scope of IPS²-PLM

"Michael Abramovici, Philip Gebus*, Jens Christian Göbel, Hoang Bao Dang"

Ruhr-Universität Bochum, Universitätstraße 150, Bochum, Germany

* Corresponding author. Tel.: +49 (0)234/32-28492; fax: +49 (0)234 32 - 14443. *E-mail address:* philip.gebus@itm.rub.de

Abstract

The Product Lifecycle Management (PLM) approach faces new challenges if transferred for Industrial Product Service Systems (IPS²). The vast amount of heterogeneous data generated throughout an IPS²'s lifecycle complicates the retrieval of required information for IPS² actors. However, these actors' risky decisions determine an IPS²'s success during use phase. Thus, an approach is needed that supports actors in finding targeted information. This paper presents a framework that utilizes semantic and text mining techniques in order to improve the information retrieval process in IPS²-PLM and to allow IPS² actors to focus on their value-adding tasks rather than spending a lot of time for finding information.

Peer-review under responsibility of the scientific committee of the 8th Product-Service Systems across Life Cycle

Keywords: IPS², Semantic Information Retrieval, IPS²-PLM, Semantic Search

1. Introduction

The lifecycles of Industrial Product Service Systems (IPS²) are shaped by a close interaction between providers and customers as well as interconnected product and service structures with strong dynamics of change during the use phase. New business models obligate providers to guarantee their IPS²s' contracted utility through a network consisting of the provider, different customers and service-partners, such as third-party service providers or suppliers [1]. Due to the vast amount and heterogeneity of the information available, it is difficult to realize a transparent, efficient and consistent information flow throughout the IPS² lifecycle in order to meet contractual obligations.

Because of the previously mentioned strong dynamic of change, especially during the use phase, a number of risky decisions made by different actors determine an IPS²'s success. Such decisions imply far-reaching consequences and therefore require an effective and fast information retrieval. As an example, an IPS² provider schedules major maintenance intervals every few weeks. If a maintenance worker locally identifies irregularities, he requires further inquiry in documents or comments, which might imply information necessary in order to decide whether unscheduled (in addition to the regular intervals), proactive maintenance is required. This valuable information is mainly hidden in various unstructured text documents, such as protocols, (maintenance) reports or comments [2]. Therefore, these documents need to be available for all involved parties in the IPS² partner network.

Product Lifecycle Management (PLM) approaches support the integration and management of product data and documents as well as engineering processes and applications [3]. Unfortunately currently available PLM solutions mainly support the single management of product, service or software-related information but not the combination of the above-mentioned [1]. Therefore, in our previous work we introduced an ontology-based PLM concept, which is suitable for IPS². Firstly because it enables a consistent information flow through the IPS²-specific heterogeneous partner networks and secondly supports the integrated management of product, service and software-related data along the whole lifecycle [4].

Commercial PLM solutions only implement traditional search engines that allow a keyword-based Boolean search in a document's metadata or at most full text search. Due to the diversity of IPS² partners, documents in IPS² provider networks differ in terms of structure and terminology used, which render a targeted information retrieval as very time-consuming or even impossible with current state of the art solutions. Thus,

Peer-review under responsibility of the scientific committee of the 8th Product-Service Systems across Life Cycle
doi:10.1016/j.procir.2016.03.083

PLM solutions deliver too many or irrelevant results on queries [2]. The vast amount of documents that are created along the whole IPS²'s lifecycle and the limited search possibilities (keyword-based) compound the aforementioned problem.

Therefore, a semantic approach is necessary that allows an intuitive, targeted and fast information retrieval within a vast amount of heterogeneous data in order to enable efficient knowledge management in IPS² partner networks.

2. Basic Requirements

Considering IPS²-specific lifecycle and data aspects in order to support actors in finding necessary information, the following groups of requirements have been identified:

A dynamic and semantically rich data model: The framework has to provide a data model that links document contents and metadata with their IPS²-PLM-specific contexts on meta data level as well as on instance level, such as concrete IPS²s, goods, services and business models. It must support heterogeneous documents due to the diverse nature of IPS² partner networks. Furthermore, it has to model the semantic relations between diverse unstructured documents and has to be dynamically expandable when new documents and instances are added.

Exploitation of semantic data: The framework has to allow actors to exploit the data model intuitively, meaning in an easy to learn and quick manner. It must enable information retrieval within the semantic (IPS² context-specific) structure as well as the content of the data provided for actors with different knowledge bases and languages.

Facilitation of targeted and outright retrieval: The framework must provide outright and targeted information retrieval results. These results have to be weighed according to a user's information needs, which he formulates within a query. Only a suitable ratio between search result precision and recall reduces the time necessary for searching information.

3. State of the Art

Traditional information retrieval mainly bases on the frequency of words in documents (text mining), whereas semantic search can be defined as a search process that exploits domain knowledge and uses formal semantics in any of the three stages of a search process: query, search process or information presentation [5,6]. Utilizing domain knowledge and semantics helps in providing targeted results to a user's specific query.

The authors in [7], for example, integrate a semantic relationship manager into the PLM approach as a framework between different coexisting software systems. They focus on semi-structured data in the beginning of life phase of products and therefore, do not consider IPS²-specific data in the use-phase or unstructured data. In [8], the authors developed the infrastructure for a semantic search engine but did not yet implement the engine itself and do not consider IPS²-specific relationships. The authors in [9] present an ontology-based context model that exploits users' context information, such as e.g. his current task, role, location or status in order to recommend and reason possibly required information. They rather present a recommendation approach than a search engine based on an ontology that was implemented in protégé and did not regard the aforementioned specifics of integrated products and services. Due to the lack of IPS² consideration in the related work, a new information retrieval approach for IPS²-PLM had to be designed.

State of the art information retrieval approaches differ on the one hand in the documents considered and on the other hand in the methods of applying semantic technologies in order to exploit domain knowledge. Current semantic technologies in the search process can be divided into the following four basic techniques: graph traversal, query expansion, spread activation and RDFS/OWL reasoning [6].

In the following, methods and scientific approaches belonging to research areas (domain knowledge representation, text mining and semantic technologies during the retrieval phase) that are necessary to develop a semantic information retrieval for IPS²-PLM are presented. Section 4 describes the precise field of application and the relations between the presented techniques.

3.1 Representing Domain Knowledge

Several methods exist to represent domain or enterprise-specific information, which differ in the way of structuring information. Glossaries, for example, are a simple collection of terms. Taxonomies classify terms by clustering them in categories. Thesaurus are also used in order to cluster terms but further express the relationships between words in form of synonyms, hypernyms or hyponyms [10]. The framework can use a thesaurus in order to overcome the difficulty of extracting information from heterogeneous documents with different terminologies.

Map representations focus on linking topics to each other and provide an overview of similar and related topics [11].

Semantic networks describe relations between concepts and terms as a graph that consists of vertices (concepts and terms) and edges (relationships) that link vertices to one another [12].

Ontologies are based on the semantic network concept. Gruber defines ontologies as "a specification of a conceptualization" [13]. Compared to semantic networks, ontologies are more strict and formal due to their axiomatic structure [13]. As the framework requires a dynamically changing data model, which does not follow strict axiomatic rules, semantic networks are preferred over ontologies for representing the relations of documents.

3.2 Text Mining

As stated before, valuable information is mainly hidden in various unstructured text documents. Documents generated during the use phase are mainly of this type. Text mining describes the process of analyzing unstructured text documents. Documents are transformed into their vector representation, where each term is stemmed and represented by its' normalized term frequency multiplied by its' inverse document frequency (measure for a term's importance) [14]. The vector space representation can be exploited to determine a document's relevancy to a given query and the similarity between two documents by calculating the cosine similarity