Computers in Industry 71 (2015) 24–34

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Review

Semantic annotation for knowledge explicitation in a product lifecycle management context: A survey

Yongxin Liao [a,b,c,*], Mario Lezoche [a,b], Hervé Panetto [a,b], Nacer Boudjlida [d,e], Eduardo Rocha Loures [c]

[a] Université de Lorraine, CRAN, UMR 7039, Boulevard des Aiguillettes, B.P. 70239, 54506 Vandoeuvre-lès-Nancy, France
[b] CNRS, CRAN, UMR 7039, France
[c] Pontifícia Universidade Católica do Paraná, PPGEPS, Rua Imaculada Conceição, 1155 Curitiba, Brazil
[d] Université de Lorraine, LORIA, UMR 7503, Boulevard des Aiguillettes, B.P. 70239, 54506 Vandoeuvre-lès-Nancy, France
[e] CNRS, INRIA, LORIA, UMR 7503, France

ARTICLE INFO

Article history:
Received 29 August 2014
Received in revised form 5 March 2015
Accepted 10 March 2015
Available online 30 March 2015

Keywords:
Semantic annotation
Knowledge explicitation
Formalization
Semantic interoperability
Product lifecycle management

ABSTRACT

Nowadays, the need for systems interoperability in or across enterprises has become more and more ubiquitous. Many research works have been carried out in the fields of information exchange, transformation, discovery and reuse. One of the main challenges in these researches is to overcome the semantic heterogeneity between enterprise applications along the life cycle of a product. As a possible solution to assist the semantic interoperability, the semantic annotation has gained many attentions and widely used in different domains. We collect a number of literature that applied semantic annotations on different objects, and classify them according to the subject being described in an enterprise architecture framework. A detailed survey, especially from the formalization aspect, is presented to identify the existing drawbacks and to point out the possible research directions.

Contents

1. Introduction

In manufacturing enterprises, the Product Lifecycle Management (PLM) approach has been considered as an essential solution for improving the product competitive ability. It aims at providing a shared platform that brings together different enterprise

* Corresponding author at: Pontifícia Universidade Católica do Paraná, PPGEPS, Rua Imaculada Conceição, 1155 Curitiba, Brazil. Tel.: +55 41 9677 0753.
E-mail addresses: yongxin.liao@pucpr.br (Y. Liao), mario.lezoche@univ-lorraine.fr (M. Lezoche), herve.panetto@univ-lorraine.fr (H. Panetto), nacer.boudjlida@loria.fr (N. Boudjlida), eduardo.loures@pucpr.br (E.R. Loures).

http://dx.doi.org/10.1016/j.compind.2015.03.005
0166-3615/

applications at each stage of a Product Life Cycle (PLC) in or across enterprises [1]. Although the main software companies are making efforts to offer a complete and integrated set of systems, most of them do not provide a coherent integration of the entire information system. This results in a kind of "tower of Babel", where each application is considered as an island in the middle of the ocean of information, managed by stakeholders along the life cycle of a product.

Semantic interoperability is the ability to ensure that the exchanged information has got the same meaning considering the point of view of both the senders and the receivers [2]. In the context of a PLM, stakeholders have to work together on the exchanged information and make decisions based on it. They have different backgrounds, heterogeneous expertise, unique knowledge, particular needs and specific practices, which over increase the difficulty to achieve semantic interoperability [3]. The mutual understanding of the semantics that is embedded inside the exchanged information is the cornerstone in the quest for semantic interoperability. Being a way to realize this enrichment, the semantic explication [4] is not only just attaching the formal and shared terms between stakeholders to make semantics explicit, but also bringing the possibility to perform the semantic reasoning for some further operations.

Several semantic annotation surveys have already been carried out with the focus on the functionality or efficiency aspect of annotation tools. However, little attention has been paid to the in-depth study and comparison of the methods behind those tools. The objective of this paper is to address this existing issue through a detailed survey on a number of semantic annotation literature, which are collected and classified from the PLM perspective. The rest of this paper is organized as follows: Section 2 presents the definitions of annotation and semantic annotation. Section 3 illustrates and compares the semantic annotation researches being applied on different objects. Section 4 identifies the existing drawbacks and proposes the possible research directions. Section 5 concludes this paper and points out the future work.

2. Annotation and semantic annotation

The Oxford dictionary defines an annotation as "*a note by way of explanation or comment added to a text or diagram*". It has special usages in different contexts. For example, in the software programming, an annotation is represented as a text comment embedded in codes to explain the program. In the mechanical drawing, an annotation is a snippet of text or symbol with specific meanings that illustrates the corresponding annotated part. In the commercial advertising, an annotation is usually used as a kind of footnote to detail some business restrictions.

In order to distinguish the semantic annotation from the other annotations, several kinds of classifications are proposed. Bechhofer et al. [5] categorized annotations into three types: the *textual annotation*, which adds notes and comments to an object; the *link annotation*, which extends the previous type of annotation by linking the object to an annotation content; the *semantic annotation*, which contains the human-readable as well as machine-readable information. Similarly, Oren et al. [6] proposed to classify annotations as: the *informal annotation*, which is expressed in an informal language and is not machine-readable; the *formal annotation*, which is machine-readable, but without any ontological terms; the *ontological annotation*, which is only composed of ontological terms that are commonly accepted and understood in a specific domain. These classifications identify two important features of a semantic annotation: (1) it is both human-readable and machine-readable, and (2) it contains a set of formal and shared terms that can exist for a community of human and/or machine agents.

Considering the essential of an ontology [7], which is a common agreement of a conceptualization of terms in a specific domain, different researchers have suggested many definitions of the semantic annotation related to an ontology. For example, Talantikite et al. [8] described it as "*a semantic annotation is referent to an ontology*". Lin [9] considered it as "*an approach to link ontologies to the original information sources*". Kiryakov et al. [10] defined it as "*a specific metadata generation and usage schema, aiming to enable new information access methods and to extend the existing ones*". To the best of our knowledge, a semantic annotation can be considered as a means to perform the semantic enrichment of "something" by using a set of well formalized and commonly agreed terms from a specific domain, such as ontologies.

In this paper, we mainly pay attention to two aspects of semantics that are made explicit through a semantic annotation: The *domain semantics*, which describes the context and the meaning of an annotated element in a specific domain; the *structure semantics*, which describes the interrelations between an annotated element and the other elements related to it. Taking into account these two aspects of semantics and the investigations that we have made in previous works [11–13], in the next section, we will discuss different semantic annotation methods inside the collected literature.

3. The investigation of semantic annotation researches

In the last decade, several surveys of semantic annotation researches have already been made with different focuses. Reeve and Han [14] presented a short survey about the classification and evaluation of six semantic annotation platforms. Uren et al. [15] reviewed and classified twenty seven semantic annotation systems according to the knowledge management requirements that they proposed. Mangold [16] presented a categorisation scheme for the classification of ten selected semantic search approaches and identified the open issues that are not addressed by those systems. Lautenbacher and Bauer [17] presented a survey to categorize and compare twenty one annotation approaches about semantic web services, grid workflows, and business process management. Hanbury [18] summarized five types of image annotation methods and then used it to analyze ten annotated image datasets. Dasiopoulou et al. [19] made a survey on eight image and seven video annotation tools, from both functionality and interoperability perspectives, to highlight the issues of the communication, sharing and reuse of produced metadata. Oliveira and Rocha [20] introduced and briefly compared nineteen semantic annotation tools to show the challenge in the quest to fully automatic annotation. Joksimovic et al. [21] presented an empirical study on three ontology-based semantic annotators to discover the issues for the future development of those examined tools.

We can find that the surveys [14–16,20,21] were mainly focusing on documents, as well as the surveys [18,19] paid major attention to images or videos. They analyzed some existing annotation tools from both the functionality aspect [15,16,18–20] and from the efficiency aspect [14,21]. A number of self-defined requirements are used as the basis to compare the semantic annotation approaches in the surveys [15,17]. Depict efforts have been made by above-mentioned surveys, at least two shortcomings need to be noted: (1) most of the surveys concerning the approaches that applied semantic annotations on texts, images, or videos, and few surveys have addressed models. The survey [17] is the only one that concerned the annotation on a specific kind of model (e.g. workflows). However, for that research, a more recent and detailed analysis is still lacking; (2) among these surveys, only three of them [14,18,20] have taken into account and generally discussed the semantic annotation methods that are embedded