Journal of Computer and System Sciences 82 (2016) 69–90

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Dynamic class hierarchy management for multi-version ontology-based personalization

Fabio Grandi

Department of Computer Science and Engineering (DISI), Alma Mater Studiorum – Università di Bologna, Viale Risorgimento, 2, I-40136 Bologna, Italy

A R T I C L E   I N F O

Article history:
Received 27 October 2014
Received in revised form 10 April 2015
Accepted 3 June 2015
Available online 10 June 2015

Keywords:
Ontology
Tree
Graph
Temporal database
Evolution
Versioning
Personalization

A B S T R A C T

We introduce a storage scheme that allows the representation and management of the evolving hierarchical structure of a multi-version ontology in a temporal relational database. The proposed scheme is aimed at supporting ontology-based personalization and temporal access to large collections of resources (data, documents, procedures etc.) stored in a dynamic environment. Whereas in previous works we considered tree-shaped ontologies only, in this work we consider ontologies with a class hierarchy structured as a general directed graph, that is also supporting multiple inheritance and intersection classes. We will also show how multi-version ontologies must be dealt with for the processing of ontology-based personalization queries.

1. Introduction

In the age of big data, when large amounts of potentially interesting and useful resources are published online day by day, the availability of semantics-aware search engines and intelligent personalization services becomes a key factor for a fruitful exploitation of such resources. In particular, the adoption of reference ontologies and their deployment for the personalization of multi-version resources has been recently proposed by several authors in the medical domain [1–4] and other application fields (e.g., e-Government [5]). The considered resources range from descriptive data to textual documents, from Web pages to the specification of processes and services. References to ontology classes are added to the computer encoding of resources (e.g., for which an XML [6] format can conveniently be used) to introduce a sort of semantic indexing of contents representing their *applicability*, *relevance* or *eligibility* with respect to ontology classes. Hence, starting from a user-supplied list of ontology classes, a suitable query engine can exploit semantic indexing to retrieve the relevant contents only and produce a *personalized version* of the desired resources.

However, in a dynamic environment, the management of this kind of semantic versioning is interleaved with temporal aspects. For example, we can choose as resources clinical guidelines [7], that is "best practices" encoding and standardizing health care procedures, in textual or executable format, and consider their personalization with respect to an ontology of diseases, patients or available hospital facilities they are applicable to [1]. Personalization will then produce guideline versions tailored to a specific use case. The fast evolution of medical knowledge and the dynamics involved in clinical practice imply the coexistence of multiple temporal versions of the clinical guidelines stored in a repository, which are continually subject to amendments and modifications. Therefore, it is crucial to reconstruct—borrowing the term from the

E-mail address: fabio.grandi@unibo.it.

http://dx.doi.org/10.1016/j.jcss.2015.06.001

legal domain—the *consolidated version* of a guideline as produced by the application of all the modifications it underwent so far, that is the form in which it currently belongs to the state-of-the-art of clinical practice and, thus, must be applied to patients today. However, also past versions are still important, not only for historical reasons: for example, a physician might be called upon to justify his/her actions for a given patient at a past time on the basis of the clinical guideline versions applicable to the pathology of patient and which were valid at that time (as well as, in the legal domain, a Court might be called to judge today on a crime committed several years ago and for which the normative framework that was in force then has to be applied).

Moreover, in a dynamic environment, the definition of domain ontologies themselves is also subject to modification and, thus, ontologies come out versioned as a consequence of updates periodically effected by domain experts and knowledge engineers, or even by standardization committees. As we showed in [8] for the legal domain (but it also happens for the medical field), personalization of a resource with respect to a past point in time must be effected by taking into account, in order to consider semantic indexing of the desired temporal version of the resource, the version of the reference ontology that was valid at the same time point. In other words, the selected resource version and the ontology version used for personalization must be *mutually temporally consistent*, in order to reconstruct the exact framework in which the resource had been utilized. Since clinical guidelines have also been recently proposed to be used as evidence of the legal standard of care in medical malpractice litigation [9], enforcement of temporal consistency is crucial to assess *a posteriori* the responsibility of physicians having followed the guidelines in the past.

Therefore, in this work we will show how temporal multi-version ontologies can be represented and maintained in a relational setting and how they can be used during the processing of a personalization query. The rest of the paper is organized as follows: in Section 2, the ontology-based personalization method proposed in [1,5] is briefly recalled; in Section 3, we present our storage scheme and manipulation primitives for temporal versioning of an ontology class hierarchy which is necessary to support such personalization method. Section 4 is devoted to personalization query processing in the presence of a multi-version ontology. Related work is discussed in Section 5 and conclusions can finally be found in Section 6.

2. A framework for ontology-based personalization

The personalization method proposed in [1,5] is based on the adoption of reference domain ontologies and the introduction of semantic indexing of resource contents with respect to ontology classes. For example, in the medical domain, reference ontologies to be used to this purpose can be derived from the ICD-10[1] international classification of diseases or from the UMLS[2] or SNOMED-CT[3] comprehensive biomedical and healthcare terminologies. Semantic indexing can then be used by personalization services to adapt generic resources to specific use cases, for example, to derive and enact individual care plans as proposed in [1–3]. Notice that, as a consequence of the information flooding we have been experiencing in recent years, personalization becomes a practical necessity when the huge availability of potentially interesting resources tends to be overwhelming.

Full ontology features, including properties, axioms, expressions and individuals, can be exploited in a processing step, which precedes the personalization process, in order to formalize the personalization context as a set of relevant ontology classes that define a specific use case. For example, in the medical domain, during this phase (called *classification phase* in [5]), a suitable reasoning facility can be used to match the medical records of a patient with the qualifying classes in an ontology of diseases. Then, such ontology classes are used as input of the personalization engine, which retrieves the data resources which are applicable, relevant or eligible with respect to such classes. The only ontology feature which is necessary for the considered personalization approach, on which [1,5,8,10,22] and this paper are focused, is the *hierarchy of classes induced by the IS-A relationship* and, thus, we do not consider in this work properties or other ontology features including the presence of instances. We initially follow the simplified assumption made in the application papers [1,5] and in the preliminary version of this work [10] that the class hierarchy underlying the ontology is tree-shaped, that is each node in the class hierarchy (but the root) has a single parent. However, in this work we will finally remove such limiting hypothesis and also consider general ontologies starting from Section 3. Owing to the tree structure, nodes can be assigned a preorder and a postorder code, corresponding to the sequence in which nodes are visited during a preorder or postorder traversal of the tree, respectively. Hence, preorder and postorder codes can be used for characterizing the descendants of a node [11,12]:

$$\text{N is a descendant of M iff M.Pre} < \text{N.Pre and N.Post} < \text{M.Post} \quad (1)$$

with obvious meaning of the used dotted notation. As we will see in Section 4, efficient testing of the descendant relationship is a key feature of personalization query processing [1].

For example, we can consider the sample ontology depicted in the left part of Fig. 1, where the corresponding preorder and postorder code of nodes can be found in the table to the right. The structure of the class hierarchy is completely defined

[1] http://www.who.int/classifications/icd/en/.
[2] http://www.nlm.nih.gov/research/umls/.
[3] http://www.ihtsdo.org/snomed-ct/.