Information Systems 52 (2015) 1–20

Contents lists available at ScienceDirect

Information Systems

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

Incorporating ontology-based semantics into conceptual modelling

Somluck La-Ongsri [a,b], John F. Roddick [a,*]

[a] School of Computer Science, Engineering and Mathematics, Flinders University, South Road, Tonsley, SA 5042, Australia
[b] School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi 11120, Thailand

ARTICLE INFO

Article history:
Received 1 April 2009
Received in revised form
17 September 2014
Accepted 9 February 2015
Recommended by M. Lenzerini
Available online 18 February 2015

Keywords:
Class hierarchy
Conceptual modelling
ER
EER
Ontology
OntoER
OntoORM
OntoUML
ORM
Semantics
UML

ABSTRACT

With the increasing complexity of applications and user needs, recent research has shifted from a data-information level to a human semantic level interaction. Research has begun to address the increasing use and development of ontologies in various applications, strongly motivated by the semantic web initiative. However, existing conceptual models are not rich enough to incorporate ontologies in one single conceptual schema. To improve this situation, it is necessary to refine modelling formalisms and make them more expressive while ensuring they remain semantically sound. We argue that conceptual modelling methodologies would be semantically richer if they were able to express the semantics of a domain that arises in concrete application scenarios. This paper investigates the incorporation of ontologies into three popular conceptual modelling methodologies, presenting the Ontological Entity-Relationship (OntoER) model, Ontological Object Role Modelling (OntoORM) and the Ontological Unified Modelling Language (OntoUML) class diagram. An extended conceptual framework for modelling ontologies and a transformation algorithm for mapping ontological constructs to relational schemata are provided so that querying the database through the conceptualisation of the database can be managed.

1. Introduction

Ontologies have been applied in a multitude of areas in computer science. The first significant growth of interest in the subject appeared in the mid 1990s which was motivated by the need to create principled representations of domain knowledge for the knowledge sharing and reuse community in the field of artificial intelligence [1,2]. A major challenge with ontologies is how to access, store and manage them. This research direction has included:

- supporting ontology-based semantic matching, querying and referential constraints in RDBMS [3–5] leading to recent advances in ontology management in DBMSs such as those introduced by Oracle,
- implementing ontology systems or tools that support ontology-based applications [6–9], such as Protégé which is an ontology and knowledge-base editor that allows the user to construct a domain ontology, customise data entry forms and enter data [10].

Another challenge driving the database community is to create better data models. These research projects have attempted to use conceptual data modelling in supporting ontologies [11,12]. For example, this has included research into methodologies for supporting database design creation and evaluation that makes use of domain-specific knowledge about an application stored in the form of domain ontologies

* Corresponding author. Tel.: +61 8 8201 5611; fax: +61 8 8201 2904.
E-mail addresses: somluck.la-ongsri@flinders.edu.au (S. La-Ongsri), John.roddick@flinders.edu.au (J.F. Roddick).

http://dx.doi.org/10.1016/j.is.2015.02.003
0306-4379/

[13,14]. In a further example, a theory of ontology was promoted that can be used to clarify the meaning of relationship constructs that are widely used to undertake conceptual modelling [15].

With the increasing complexity of actual application scenarios and user needs, there is a requirement to shift from the data and information level to the human semantic level interaction. Consequently, semantic representation becomes important and to maximise the level of semantics requires that these representations become increasingly explicit. Humans learn to deal with the ambiguity of language by understanding the context in which terms are used. Data rich systems can emulate this by referencing data through structures such as *ontologies* that represent terms and their interrelationships [3].

However, the conceptual modelling for ontologies and its transformation into relational database schema has not been widely investigated. This paper supports the argument that conceptual modelling methodologies must be expanded to facilitate ontologies, including the reuse of existing ontologies, to enrich the semantic expressiveness of the data model. We suggest extensions to high-level conceptual models to represent the relationship between ontologies and the underlying conceptual schema. In addition, our approach caters for the ability to query the data in the context of its associated ontologies in the same way as querying simple relational data.

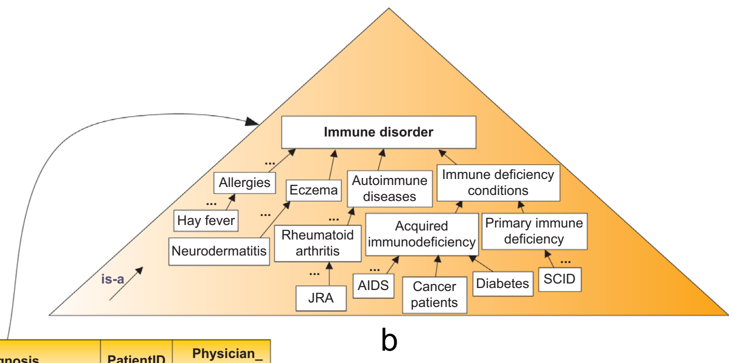
Consider a particular example of a medical database application that requires ontologies, specifically the knowledge associated with a hierarchical domain. For a simple application of a patient's visit to a doctor, a relational table with a `Diagnosis` attribute as shown in Fig. 1(a) can describe the diagnosis of a disease as identified by a physician at a visit date. The diagnosis attribute's domain is a hierarchical structure which can be represented as a diagnosis ontology as shown in Fig. 1(b).

Consider the following question: *how can those patients that have been diagnosed with immune deficiency conditions be identified?* Current practice would be by identifying those patients with AIDS and so on. This response occurs as humans have the ability to combine data with the domain knowledge that AIDS is a type of immune deficiency condition, and in many instances this connection is made automatically at a subconscious level. However, the fact that AIDS is a type of immune deficiency condition is not explicitly represented in the data as shown in the table in Fig. 1(a), but belongs to the domain knowledge of diagnosis ontology as shown in Fig. 1(b).

Within the medical database, the relational DBMS allows us to query on the attributes `Date`, `Diagnosis`, `PatientID` and `Physician_LicenceNo`. This expressive power of traditional queries is relational complete. However, for querying such attributes whose domain, for example, is related to the ontology hierarchy (by relationships such as `is_a`), the traditional query is limited such as in the situation where it is necessary to identify those patients who are diagnosed with immune deficiency conditions. In this scenario, a query on a conventional database application could be constructed using the equality operator (=) as follows:

```
SELECT    PatientID
FROM      visit
WHERE     Diagnosis = 'Immune deficiency conditions';
```

This traditional SQL query will fail to return any results that semantically satisfied the query condition since none of the values of `Diagnosis` in the table will match the text 'Immune deficiency conditions'. In other words, the domain knowledge required to answer such queries is not present in the relational table in Fig. 1(a). To provide



VISIT

Date	Diagnosis	PatientID	Physician_LicenceNo
4/01/2008	Hay fever	0001	A1515
15/02/2008	Neurodermatitis	0002	B0092
7/11/2007	Rheumatoid arthritis	0003	C0185
6/12/2007	AIDS	0004	D1006
19/12/2007	SCID	0005	E0005
14/01/2008	Diabetes	0006	F0006

a

Fig. 1. A portion of a medical database (a) the base table of patients visiting a doctor, and (b) a diagnosis ontology.