Multi-Layered Semantic Data Models

László Kovács
University of Miskolc, Hungary

Tanja Sieber
University of Miskolc, Hungary

INTRODUCTION

One of the basic terms in information engineering is data. In our approach, data item is defined as representation of an information atom stored in digital computers. Although an information atom can be considered as a subject-predicate-value triplet (Lassila, 1999), data is usually given only with its value representation. This fact can lead to definitions where data is just numbers, words or pictures without context. For example in (WO, 2007), data is given as information in numerical form that can be digitally transmitted or processed. It is interesting that *we can often recognize that the term 'data' is used without any exact terminological definition with the effect that the term often remains confusing, sometimes even contradicting the definitions of the term presented. Sieber and Kammerer (2006) introduce a new interpretation of data containing several levels. The lowest level belongs to data instances that describe the form and appearance of symbols. The intermediate level is the level of representatives which includes the applied encoding system. The highest level is related to the meaning with context description. All three levels are needed* to get to know the information atom. For example the symbol '36' in a database determines only the value and representation system, but not the meaning. To cover the whole information atom, the database should store some additional data items to describe the original data. The main purpose of semantic data models is to describe both context and the main structure of data items in the problem area. These additional data items are called metadata. It is important to see that:

- metadata are data,
- metadata are relative, and
- metadata describe data.

Metadata constitute a basis for bringing together data that are related in terms of content, and for processing them further. They can be understood as a pre-requisite for intelligent and efficient administration and processing, and not least as a focused, formal means of providing relevant data.

BACKGROUND

In data management systems, the context of a value is usually defined with the help of a storage structure. An identification name (a text value) is assigned to each position of the structure. The description of storage (structure, naming and constraints) is called schema. A big problem of structural data modeling is that it can not provide all the information needed to understand the full context of the data. For example, a relational schema

RT (NM INT, KNEV CHAR(20), RU DATE)

alone is not enough to capture the meaning of the stored data items.

The main building blocks to describe the context in semantic data models (SDM) are concepts and relationships. The first widely known structure oriented semantic models in database design are the Entity-Relationship (ER) model (Chen, 1976) and the EER (Thalheim, 2000) model. The ER model consists of three basic elements: entity (concept), relationship and attribute. The attributes are considered as structure elements of the entities, one attribute may belong to only one entity. The EER model is the extension of the ER model with IS_A and HAS_A relationships. Some other extensions are SIM, IFO and RM/T. One of the main drawbacks of structure oriented SDM is the limitations of expressive power.

Later, models like UML or ODL (Catell, 1997) were developed to cover the missing object oriented elements. In the case of ODL, a class description can contain the following elements: attributes, methods, inheritance parameters, visibility, relationships and integrity rules. These models provide a powerful complexity for software engineering but they are not very flexible to describe data models of higher abstraction.

Global investigations were focused on the SDM with simpler and more universal elements. The most widely known high level semantic models are semantic networks and ontology models. A semantic network is represented with a directed graph where the vertices are the concepts and the edges are the relationships. The main differences between ontology models and the traditional SDM are in the followings: there is no fixed structural hierarchy among the concepts, flexible relationships, independence from application domain, structure is mapped into a logical formula, it can be related to an inference engine. It is widely assumed that anything at a high level of information processing must be based on ontology (Sloman, 2003). Further details can be found on current applications of ontology among others in (Taniar, 2006).

One of the first languages for ontology is RDF (Lassila, 1999). RDF is used to describe the concepts in a neutral, machine-readable format. According to the specification, the basic language elements are resources, literals and statements. There are two types of resources: entity resources and properties. A statement is a triplet (p,s,o), where p is a property, s is a resource and o is either a literal or a resource. In another approach, p is called predicate, s is the subject and o is the object in the statement. As it can be seen, a statement corresponds to an information atom.

A pioneer representative of the next generation of languages is OWL (Bechhofer, 2004) which can be considered as an extension of RDF, that contains extra elements to describe among others typing, property characteristics, cardinality and behavioral properties. The OWL-DL language is based on Description Logic that describes the structural relationships of the domain in a logic language, which enables automatic reasoning and constraint checking in the system. The applied logic language is based on first-order predicate logic. The most widely used products related to OWL are Protégé, Pellet and KAON2.

MULTI-LAYERED SEMANTIC MODELS

Multi-Layered Schemas

In the case of systems with complex functionality, one way to reduce complexity is to build up a modular system. Modularization is a successful concept in all engineering areas. Modularization can be vertical or horizontal. Vertical modularization is called layering. The basic properties of a layered system are the followings:

- the elements are assigned to clusters (called layers);
- there exists a hierarchical relationship between the clusters;
- the relationships within the clusters differ from the relationships between the clusters;
- the clusters cooperate with each other in the role of a client or of a server.

Every layer offers a set of functionality where the functions are built upon the services of the underlying layers. In the case of a multi-layered system, the implementation can gain in cost reduction compared with a single-layer structure. Layering means modularization from the viewpoint of implementation and it has the following qualitative and quantitative benefits (Knoerschild, 2003):

- encapsulation (the layers are in great part self-contained, consistency) ,
- independence,
- flexibility (the layers can be replaced without affecting the other layers),
- cost reduction (simplicity in testing and in design, reusability).

The layered structure is a common technology nowadays among others in networking (Hnatyshin, 2007), image processing (Sunitha, 2007), process control (Zender, 2007) and software development (Kreku, 2006).

Related Content

A Survey of Fog Computing-Based Healthcare Big Data Analytics and Its Security
Rojalina Priyadarshini, Rabindra Kumar Barik, Harish Chandra Dubeyand Brojo Kishore Mishra (2021).
International Journal of Ambient Computing and Intelligence (pp. 53-72).
www.irma-international.org/article/a-survey-of-fog-computing-based-healthcare-big-data-analytics-and-its-security/275758

Process Model for Content Extraction from Weblogs
Andreas Schieberand Andreas Hilbert (2014). *International Journal of Intelligent Information Technologies (pp. 20-36).*
www.irma-international.org/article/process-model-for-content-extraction-from-weblogs/114957

The Undervaluation of Family Businesses: The Explanatory Power of Governance Mechanisms
Narjess Skhiri Ep Hellara (2024). *Social and Ethical Implications of AI in Finance for Sustainability (pp. 247-267).*
www.irma-international.org/chapter/the-undervaluation-of-family-businesses/345122

Automatic Fuzzy Parameter Selection in Dynamic Fuzzy Voter for Safety Critical Systems
PhaniKumar Singamsettyand SeethaRamaiah Panchumarthy (2012). *International Journal of Fuzzy System Applications (pp. 68-90).*
www.irma-international.org/article/automatic-fuzzy-parameter-selection-dynamic/66104

Empowering Faculty Vitality and Mitigating Burnout Through Generative AI in Higher Education: Reimagining Learning Environments With Generative AI
Stacy Ybarra (2024). *Transforming Education With Generative AI: Prompt Engineering and Synthetic Content Creation (pp. 281-308).*
www.irma-international.org/chapter/empowering-faculty-vitality-and-mitigating-burnout-through-generative-ai-in-higher-education/338542