

Development of an Ontology-Anchored Data Warehouse Meta-Model

Jyoti Kamal, PhD¹; Tara Borlowsky, MA¹; Philip R.O. Payne, PhD^{1,2}

¹The Ohio State University Medical Center, Information Warehouse, Columbus, OH

²The Ohio State University, Department of Biomedical Informatics, Columbus, OH

Abstract

Data warehouses must provide a flexible data model that is integrated with knowledge and metadata describing their components and contents. To provide for advanced query functionality at The Ohio State University Medical Center (OSUMC), we have developed an abstraction layer, or meta-model for our existing Information Warehouse (IW) in order to conceptually and semantically describe and classify its structure and contents using the UMLS.

Introduction

Large-scale databases that evolve over time and interact with external systems must be flexible, maintain conceptual simplicity, and be designed such that they contain an integrated data model describing knowledge about its dynamic structure and contents¹. Although a limited number of reports have described the utilization of domain-specific metadata for standardizing and linking data from heterogeneous sources², to the best of our knowledge, such a methodology has never been applied to an existing clinical data warehouse. The Information Warehouse (IW) at OSUMC stores multi-dimensional data from multiple information systems utilized within the institution to allow researchers, clinicians and administrators to view and analyze integrated data sets. As a proof-of-concept to ease and improve upon existing query mechanisms for the IW, we have developed an abstraction layer, which utilizes several standard vocabularies (e.g., SNOMED-CT, NCI Thesaurus) and their semantic relationships, to conduct class-based reasoning over the current data model (Figure 1).

Methods

Core Tables

The components (i.e., datamarts, schemas, tables, columns) of the existing IW data model were automatically aggregated and manually inspected to ensure the existence of accurate descriptions for each and note which columns contain coded data. This information was subsequently utilized to generate the core tables of the meta-model, which describe all IW data model components and their inter-relationships. In addition, a data dictionary was derived from the information contained in all of the coded columns.

Ontology-based Knowledge-base

The UMLS Knowledge Source Server was then employed to map each unique column to all relevant concept unique identifiers (CUIs). Each column was

assigned codes corresponding to its data type (e.g., date, code) and class (e.g., diagnosis, procedure), as well as based upon the column documentation as described in the associated meta-model core tables. Each UMLS CUI was subsequently mapped to one of the source vocabularies (i.e., SNOMED-CT, NCI Thesaurus, ICD9CM, LOINC) maintained within the OSUMC instance of the Apelon Distributed Terminology System (DTS)³, which was then used to generate an *ontology-anchored representation* of the IW data model.

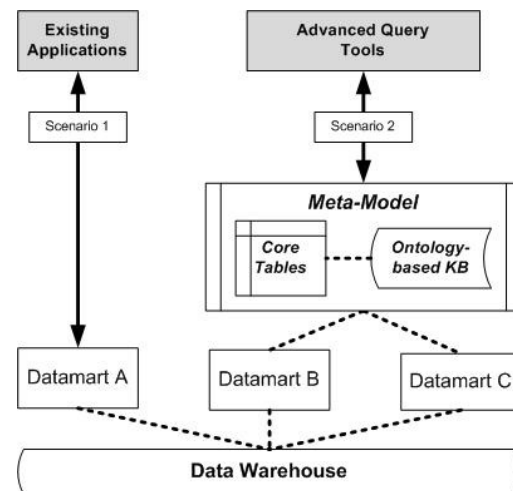


Figure 1. Insertion of ontology-anchored abstraction layer (meta-model) to support advanced query functionality over existing data warehouse.

Discussion

The implementation of such a meta-model within the OSUMC IW has many potential benefits, including more efficient responses to data requests, easier navigation and understanding of IW content by non-expert users, and ability to develop applications utilizing generic class-based query structures.

References

1. McClatchey R, Kovacs Z, Estrella F, et al. The role of meta-objects and self-description in an engineering data warehouse. International Database Engineering and Applications Symp. 1999: 342.
2. Chong Q, Marwadi A, Supekar K & Lee Y. Ontology based metadata management in medical domains. J of Res and Prac in Inform Tech. 2003; 35(2): 139-54.
3. Apelon DTS, <http://apelon-dts.sourceforge.net/>, Last Accessed: 2007, July 30.