On an EAV Based Approach to Designing of Medical Data Model for Mobile HealthCare Service

Alexander Borodin, Yulia Zavyalova

Department of Computer Science
Petrozavodsk State University
Petrozavodsk, Russia
Email: {aborod, yzavyalo}@cs.petrsu.ru

***Abstract*—As a consequence of the advances in technologies of wearable devices, the amount of data received from different kinds of sensors tend to grow significantly. This is also the case with medical data. In most cases, this highly diverse medical data has multi-dimensional and sparse structure. There are generic XML and RDF formats to describe this data. Moreover, RDF is a key technology for semantic applications. Nevertheless, in real-world high-loaded services, relational databases are used for storing large volumes of attribute-value pairs for the purposes of better performance. This paper describes the approach to database schema design for a service of continuous monitoring of health parameters. The proposed approach is based on Entity-Attribute-Value model widely used in medical databases.**

***Keywords*–m-Health; Personalized Services; Entity-Attribute-Value.**

I. INTRODUCTION

Currently, life expectancy among the elderly tends to improve. It leads to a significant growth in the number of patients with chronic diseases and it increases health care spending. Remote monitoring of chronically ill patients is the promising way of bounding the healthcare costs growth.

In the Park of Innovative Technologies of Petrozavodsk State University, two systems for the analysis of remotely harvested health data processing are being developed. The Research Platform for Building Medical Diagnostic Services is designed for processing medical diagnostic data [1]. This platform relies on Continuity of Care Record (CCR) standard and is intended for research purposes.

The aim of second project, CardiaCare is to provide a system for continuous monitoring of heart function [2]. The service concept developed in this project is more practical oriented and should be implemented in collaboration with an Emergency Hospital and with assistance of Ministry of Health of Karelia.

With the development of technologies of a mobile health-care and, in particular, of wearable sensors, the amounts of data received from different kinds of sensors grew significantly. This highly diverse diagnostic information is stored in so-called Clinical Study Data Management Systems (CSDMSs) for further analysis and decision making. One of the vital features of CSDMSs is the ability to encompass hundreds of new clinical parameters with no need of database schema updates during the life cycle of the system [3].

Both projects mentioned above use a relational databases as the CSDMS backend. However, in relational databases, a description of the kinds of data that the database stores — the attributes – is recorded implicitly in the schema in the form of the structure of tables and relationships between them. Nevertheless, in rapidly evolving circumstances related to the development of new sensor hardware, frequent schema changes will be required. Redesign of the schema will, in turn, lead to the need of reimplementing almost all components of the system.

Under these circumstances, the so-called Entity-Attribute-Value (EAV) database design is useful. Nevertheless, the EAV design has both advantages and drawbacks.

In this paper, we discuss properties of EAV-modeled database aimed at the remote monitoring of health parameters and present design decisions that have been implemented in CardiaCare service backend.

The rest of the paper is organized as follows. Section II presents an overview of the CardiaCare service and the main requirements of the storage are justified based on the architecture of the service. Section III recalls the properties of EAV model. Section IV introduces our design decisions. Section V summarizes the results of this paper.

II. OVERVIEW OF CARDIACARE SERVICE

CardiaCare is a mobile service aimed at the continuous monitoring of heart function, detection of several kinds of arrhytmias and risk factors assessment based on the joint analysis of electrocardiogram recordings, auxiliary conditions, parameters of the environment (e.g., outside temperature) and concomitant data (e.g., individual notes on stairs tests or indisposition cases).

Individual medical data is measured by the network of personal devices that are equipped with medical sensors and connectivity modules. Recordings of health parameters are sent wirelessly to the smartphone. The patient is also able to input optional notes manually. A CardiaCare mobile app provides a simple analysis of ECG recordings and arrhytmia detection, risk factors assessment and in case of emergency situation sends an SMS to a specified mobile number to inform relatives or doctor. The application provides a simple feedback to the patient, e.g., advises to slow down physical activity or to visit a doctor.

The harvested data are sent to the server for further analysis on resource-intensive algorithms and storing. On the servers side, alarms can also be generated. Workplace for a doctor is also provided.

The high-level architecture of CardiaCare service is presented in Figure 1.

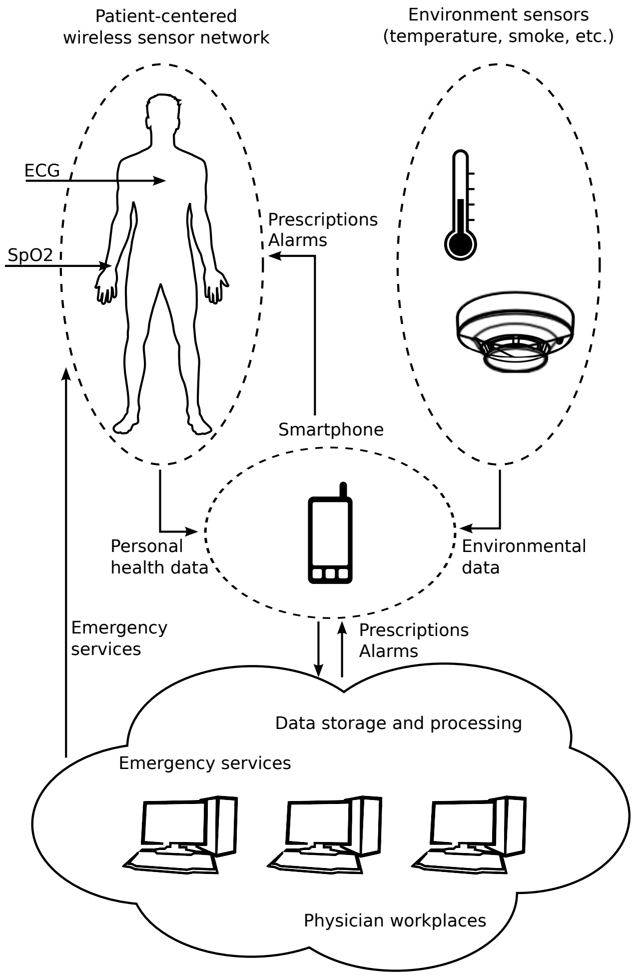

Figure 1. Using attribute dictionaries

For the purpose of improving the speed of emergency help, we propose the emergency service in terms of smart spaces paradigm on top of the CardiaCare service.

This service provides the healthcare personnel (ambulance services, volunteers, etc.) with the information of the emergency cases and location of the patients. The smartphones of participants are equipped with a software agent that has access to the risk factors and alarms generated by CardiaCare. In terms of smart spaces these agents are referred to as knowledge processors (KPs). Collecting and exchanging of the alarms and location of the participants is managed by semantic information broker (SIB). The emergency service provides the possibility of finding the closest care producer and assessment of the amount of time needed to reach the location of care receiver.

The architecture of the emergency service is presented in Figure 2.

From the description of the architecture follows that the number of sensors and other tests could vary during the life cycle of the service. Nevertheless, the design of the medical data storage should provide a possibility of flexible adaptation to these circumstances.

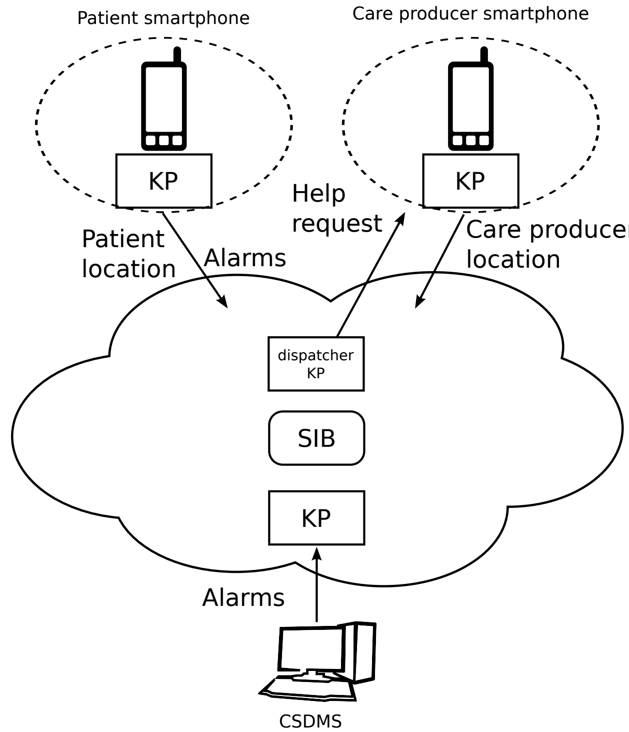

Figure 2. Using attribute dictionaries

III. Entity-Attribute-Value design

In relational databases attributes usually are represented by table columns, one column per attribute. Nevertheless, this model is suitable when modeled entities have a fixed number of attributes and most or all of them have values for any given instance. If entities have potentially large number of attributes and, for a given instance, only a few of them are non-empty (applicable or known) then representation of attributes with columns leads to very sparse tables.

This situation is distinctive for clinical databases, when for any given patient only few vital parameters are actually recorded from hundreds of available ones.

A standard way of representing arbitrary information about some object is a set of attribute-value pairs, which become triples with the entity. There are generic formats to describe such triples:

- Extensible Markup Language (XML) provides a tool of describing complex hierarchical structures;
- Resource Description Format (RDF) operates with object-attribute-value triplets.

The attribute-value representation of the information can also be appropriate within a relational database.

The basic principle of EAV design is in the way of attributes representation. Unlike column-based attributes in relational model, in EAV attributes are row-modeled, one fact about entity per row [3].

EAV approach trades off the simplicity of physical representation and complexity of mapping to logical one and querying data. Search inefficiency is one of the drawbacks of the EAV approach. To eliminate this inefficiency several optimizations have been proposed [3] [4].

 ISBN: 978-1-61208-418-3

Commonly used optimizations can be described with the following procedure illustrated in (Figure 3).

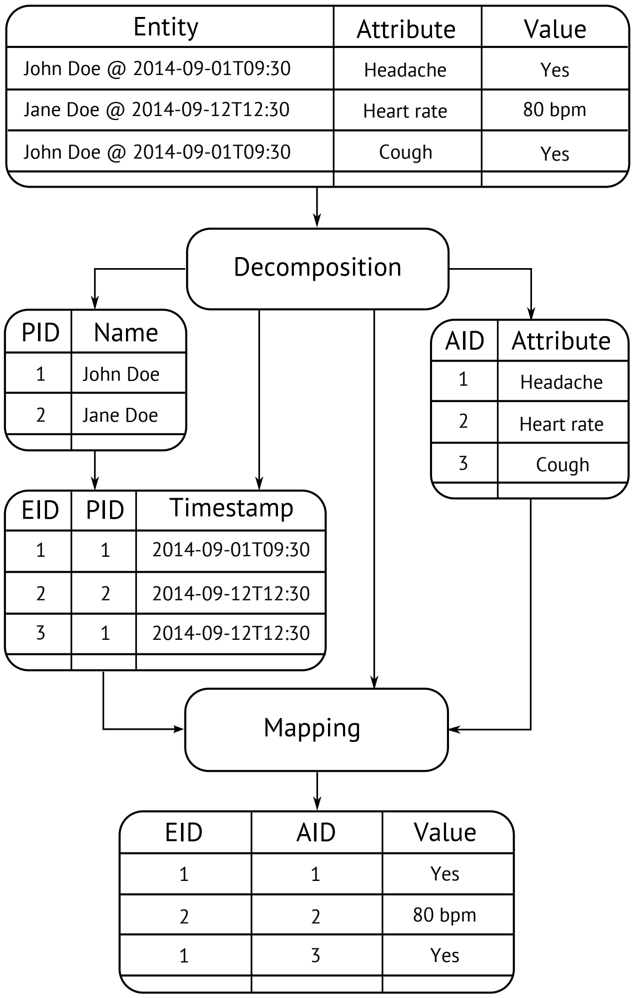

Figure 3. Using attribute dictionaries

In the first step, entities and attribute names are moved to separate dictionaries. Therefore, they are mapped to integer identifiers and the redundancy of duplicating attribute names is eliminated. In a second step, an auxiliary table is constructed to map values to entities and attributes.

Other optimizations propose additional indexes constructing, using, for example, binary representations of attribute and value, which are concatenated. In this solution, an index is constructed automatically sorting rows by attribute. It makes the model more optimized for "read" operations than original EAV and it fits the requirements of most of CSDMSs, since, in clinical trials, we write to the database once, but the data is used potentially several times. Nevertheless, in the general case, this approach is not applicable.

In generic EAV attribute values should be of the same data type. Often, these are just strings. This produces several difficulties when constructing complex queries. There are two options to avoid this inefficiency.

The first option is to use several columns for different data types (Figure 4). Only one of the column values is allowed not to be null.

The second option is to store attributes of different types in different tables (Figure 5). This approach eliminates the redundancy of the first one, but the queries should be done over all these tables.

VID	EID	AID	ValueStr	ValueInt	ValueBool
1	1	1	NULL	NULL	Yes
2	2	2	NULL	80	NULL
3	1	3	NULL	NULL	Yes

Figure 4. One table for all attribute values

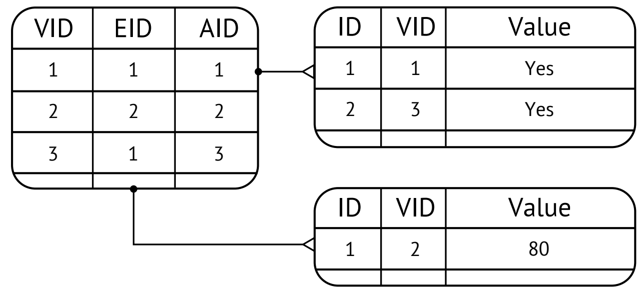

Figure 5. Separate tables for attributes of different types

The second option leads to more complex queries and increases the processing delays. Nevertheless, it helps to get rid of the numerous empty cells. Hence, the choice should be made depending of the nature of the data.

IV. EAV Approach in CardiaCare Database

A. *Overview of CardiaCare schema*

CardiaCare data model has hybrid design. Part of the tables have conventional relational column-based structure. Other tables are row-modeled and introduce the EAV approach.

Recordings of clinical trials heavily rely on predefined values of attributes that are organized as dictionaries. CardiaCare model supports compound attributes. Thus, an entity (e.g., medical trial) can be described not only by a set of attribute-value pairs, but also by a tree of attributes. Attribute dictionaries may also have hierarchical structure, as shown in Figure 6. As an example, one can consider some geo points dictionary with two-level (country-city) hierarchy.

There are two types of entities in CardiaCare data model:

- profiles of patients may be extended with arbitrary fields from contact details to billing information;
- clinical trials may be extended with arbitrary parameters and new types of trials may be incorporated as well.

Clinical trials can be considered not only as a hierarchy of parameters describing the event, but also as a form that consists of a hierarchy of fields to fill in to describe the event.

Attributes are described by metadata tables, as shown in Figure 7. The list of all applicable attributes is stored in 'attribute' table. The names of attributes are listed here and their belonging to the entities is defined. Elsewhere, a machine-generated identifier is used to refer to the attribute.

 ISBN: 978-1-61208-418-3

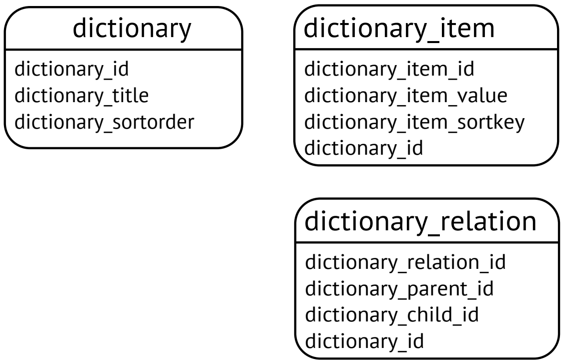

Figure 6. Hirarchical dictionaries in CardiaCare data model

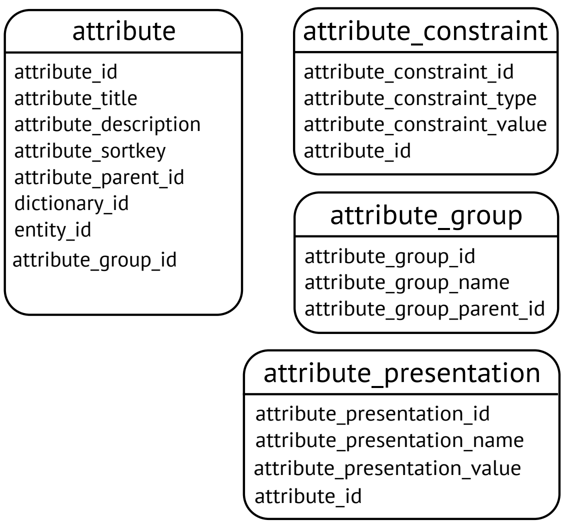

Figure 7. Attribute representation in CardiaCare data model

There are several other metadata tables, with the important ones being listed below.

- The 'attribute_constraint' table contains various metadata used for validation purposes, including data type checks, minimal and maximal allowed length of string values, allowed range of integer values, regular expressions, etc.
- The 'attribute_group' table contains attribute grouping for presentation to the user. Attribute groups can be nested.
- The 'attribute_presentation' table contains visual parameters of form fields for input forms and reports.

The aim of this auxiliary metadata is to provide an opportunity to construct user interfaces and REST APIs automatically.

V. CONCLUSION

Data model for CardiaCare service was conducted based on open schema design principles. This will allow to extend a set of supported medical measurements without the need of making changes to the database schema.

The schema was designed taking into account a possibility of automatic user and machine interface construction. Development of such framework is the purpose of the next stage of this research.

ACKNOWLEDGMENT

This research is financially supported by the Ministry of Education and Science of the Russian Federation: project # 14.574.21.0060 (RFMEFI57414X0060) of Federal Target Program “Research and development on priority directions of scientific-technological complex of Russia for 2014–2020”

REFERENCES

[1] Y. Apanasik, I. Shabalina, and L. Kuznetsova, “The Research Platform for the Medical Diagnostic Services Building,” in Proceedings of the 14^{th} Conference of Finnish-Russian University Cooperation in Telecommunications Program, Helsinki, Finland. 11.11.2013-15.11.2013, 2013, pp. 9–15.

[2] A. Borodin, A. Pogorelov, and Y. Zavyalova, “The Cross-platform Application for Arrhythmia Detection,” in Proceedings of the 12^{th} Conference of Finnish-Russian University Cooperation in Telecommunications Program, Tampere, Finland. 5.11.2012-9.11.2012, 2012, pp. 26–30.

[3] V. Dinu and P. M. Nadkarni, “Guidelines for the effective use of entity-attribute-value modeling for biomedical databases.” I. J. Medical Informatics, vol. 76, no. 11-12, 2007, pp. 769–779.

[4] R. Paul and A. S. M. L. Hoque, “Search efficient representation of healthcare data based on the hl7 rim.” Journal of Computers, vol. 5, no. 12, 2010, pp. 1810–1818.

