Integration of Computer Systems with the Use of Ontology

Ilya Semerhanov, German Vargin, Dmitriy Muromtsev
Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics
Saint-Petersburg, Russia
mouromtsev@mail.ifmo.ru

Abstract

As the title implies the article describes a new approach in system integration. Authors discuss problems of integration between complex computer systems. They give an explanation of term interoperability of computer systems and propose to use integration method based on ontology. Authors also mention that they are currently developing software application with this method. This article will be of interest to those who are researching new solutions for data and system integration.

Index Terms: System integration, Semantic web, Ontology, RDF, SSOA.

I. Introduction

Under the integration of data in computer systems understood a solution, which provides a single, unified interface to a set of independent heterogeneous data sources.

Development of methods for the integration of data resources is one of the most pressing problems in the field of computer systems. In recent years it attracts more and more attention. However, the problem of data integration is not a new one. The first steps in this area were made in the mid-70's, when the development of distributed database systems began and a clearer idea of the layered architecture of database systems of data models was formed.

It thus was primarily to support the global scheme for the collection of local databases that operate in different network nodes under DBMS that support the same or, in general, different data models. Later, a little bit more generic form of this problem was related to the creation of multidatabases, federated databases, data warehousing, various repositories of data resources, as well as web-based applications. In recent years, problems of integration of heterogeneous data began to play a key role in widely spread development of digital libraries, and there is also the problem of the integration of text data resources from various independent sources [1].

The problem of data integration is extremely diverse and complex. The complexity and nature of the methods used to solve it are strongly dependent on the level of integration that is necessary, the individual data sources properties, and the properties of the entire set of sources in general.

This article deals with the development of a solution, based on metadata and ontology, for integration of computer systems that operate in the same domain.

II. MAIN PART

The main problem of integration is that the computer systems are usually complex software systems, consisting of a large number of information units, entities, attributes, relationships. In order to exchange data between systems there is a need to transfer not only the individual values of specific variables, but also a large amount of related information. This information must be imported in the entities of other system. Also in this integration process data type conversion must be performed.

In the integration process of computer systems a key role is played by the property, called interoperability. Interoperability is the ability of diverse computer systems to work together. For example this interaction can manifest itself in data exchange, replication of databases, distributed execution of queries. When there is a need to connect different computer systems in one domain, there is always a need to ensure interoperability of these systems.

This need to ensure interoperability is the fundamental problem of system integration. This problem is relevant to both legacy systems, when they are connecting to the newly created, and also for absolutely new data storages, which should be possible to integrate with other systems in the future.

Interoperability has two aspects: syntactic and semantic. The syntactic aspect of interoperability of systems is the ability to manage structural coordination of systems entities. If two or more systems are capable of communicating and exchanging data in any way, they are using syntactic interoperability. Semantic aspect means the possibility of establishing correspondence between meanings of units of computer systems. Beyond the ability of two or more computer systems to exchange information, semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems [2].

Existing methods of achieving interoperability, mainly focused on its syntactic aspects, that is, aimed at harmonizing and converting of data structures to standard formats. Currently there are no generic approaches that aim at interoperability in the semantic aspect. All available solutions for semantic interoperability are only data storage specific, and they provide just manual mapping between entities. Existing methods and tools for integration of computer systems, based on the manual construction of the rules for data exchange between computer systems and implementation of appropriate applications adapters are not enough now for the newly developed complex systems and infrastructures.

Currently a new approach, that use metadata for information resources description, is actively developed. But for now, when it's used, it usually only provides a description of the structure of information units, rather than analysis of the meaning and purpose. Based only on these descriptions it is impossible to correctly calculate dependencies between informational units, their links in different systems that work in one domain. To form the conclusions it is necessary to operate not only with the structural characteristics of objects in computer systems, but also to evaluate their meaning.

Authors in this work are trying to find a generic solution to the problem of interoperability, using description of metadata of computer systems and implementing mapping for entities and relations in all systems in terms of specific domain ontology.

In any system, there are two general categories of perception, such as the objects, which create the structure of the system and the relationship between these objects that characterize the state of the system. In ontology, term relationship clearly describes dependencies between system objects in the real world. Ontological model represents the most important statements in the subject domain. Additionally, this model helps to describe the behavior of objects and the corresponding change in the relationship between them, that is, the behavior of the system. Thus, an ontology is a data dictionary, which includes terminology and system behavior model. As each conceptual domain model is a subset of the ontology, the problem of integration of computer systems is reduced to the problem of mapping between the models, in terms of ontology [3].

At the moment, authors started the development of a software tool for the integration of computer systems based on ontology. The solution is still in a design stage, but when the development will be finished, this tool will provide an agile approach for the integration. Description of the ontology, as well as metadata will be based on XML technologies and models of RDF. Resource Description Framework (RDF) was developed to solve problems related to the description of semantics. The central concept in RDF is the data model. This is a set of facts and semantic relations between them. RDF is an abstract model that provides a way to partition knowledge into discrete parts. Statement about a resource has the form "subject - predicate - object" and called the triple. To refer to the subject, predicate and object unified resource identifier is used in RDF (URI). The set of RDF-statements form a directed graph in which the vertices are the subjects and objects, and the edges are labeled by predicates.

The software solution will use RDF-statements based on information extracted from the relational database of computer systems, and the domain ontology. To do this, first of all data schema structure in XML: format will be extracted from the computer systems to achieve syntactic interoperability. For the implementation of semantic interoperability it will use a conceptual domain model. This model will be created for each information system using ontology and described in RDF statements. Ontology contains a dictionary of concepts and keeps the overall relations between these concepts. Then, a general model for the interaction of two computer systems is created on the basis of conceptual models and ontology. Figure 1 shows a conceptual diagram of the proposed solutions.

Application will be based on Jena framework, which is an open source Semantic Web framework for Java. It supports triple stores and also drivers for a number of different database engines, including MySQL, PostgreSQL, Oracle. Also reasoned Pellet: OWL 2 will be used in the solution. It incorporates optimizations for nominals, conjunctive query answering, and incremental reasoning. To interact with all the computer systems the proposed software will use SSOA (Semantic service-oriented architecture) [4]. SOA relies on common Web Service technologies which allow interoperability by relying on standards like UDDI, WSDL, SOAP. This means that applications based on different platforms and written in different languages can easily exchange data over the network using web services. Since HTTP is supported by all web servers and browsers, SOAP messages can be sent between applications, regardless of platform and language development. Using ontologies and semantics, services can be unambiguously described, from data, functionality and behavior point of view. These semantics descriptions can be used in operations like discovery, selection, composition or aggregation of services to provide meaningful functionality through a truly Service Oriented Architecture.

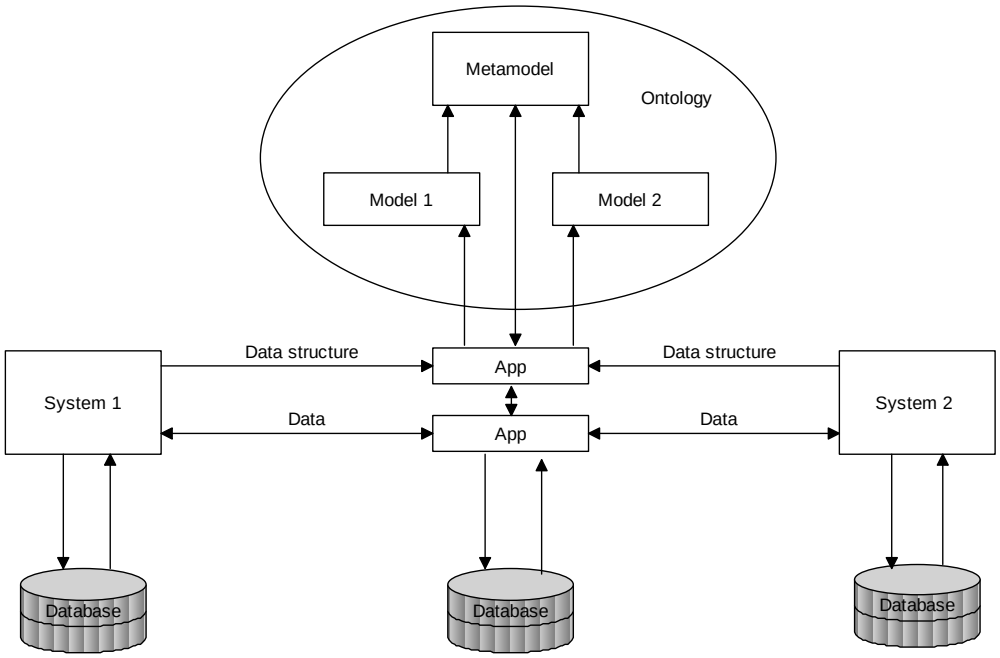

Fig. 1. Conceptual diagram of solution

III. CONCLUSION

The problem of integration of computer systems has a great importance in the modern world. The solution of this problem is a key to successful automation of business processes in organizations, which develop system infrastructure. With the successful implementation of the proposed solutions, authors will present a method and a tool that can include in a single information field of given ontology, different computer systems, as well as provide the opportunity to interact with other systems in the domain.

REFERENCES

[1] Kalinichenko L.A., Briukhov D.O., Skvortsov N.A., Zakharov V.N. “Infrastructure of the subject mediating environment aiming at semantic interoperability of heterogeneous digital library collections.” *Second Scientific Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections"*, Protvino, 2000, c. 78-90.

[2] Mikhaylov I.S “Information Systems Metamodels Developing for Providing Structural and Semantic Interoperability”, *International Conference on Dependability of Computer Systems DepCoS - RELCOMEX* 2007, Szklarska Poreba, Poland, 14-16 June, 2007 pp 59 – 64.

[3] Andrea D' Andrea, Franco Niccolucci “Mapping, Embedding and Extending: Pathways to Semantic Interoperability the Case of Numismatic Collections”, *Semantic Interoperability in the European Digital Library*, Proceedings, Tenerife, 2008.

[4] Haller A., Gomez J.M., Bussler C., “*Exposing Semantic Web Service principles in SOA to solve EAI scenarios*”, May 10–14, 2005, Chiba, Japan.