

# **Improvement of Semantic Interoperability based on Metadata Registry in a Digital Library Environment**

Doo-Kwon Baik  
Software System Lab  
Department of Computer Science, Korea University  
5-Ga, Anam-Dong, SungBuk-Gu  
Seoul, Korea  
baik@software.korea.ac.kr

## **1. Introduction**

Digital library is an application fields that federation of index databases and local databases is essential. To integrate databases, we should surmount semantic heterogeneity of database. We think that there are two level of semantic heterogeneity. The first level is heterogeneity of schema (data element), and the second is that of terminology (data value). In index databases, the schema level heterogeneity occurs owing to various metadata encoding formats (such as Dublin core, USMARC, KOMARC). And, local databases use non-standardized data element names to construct the databases. In Addition, we must consider semantic heterogeneity of terminology. An example of semantic heterogeneity is the multi-lingual problem. Because of semantic heterogeneity, it's difficult to share and exchange information among information systems.

The objective of this research is to improve semantic interoperability of index databases and local databases in a digital library environment. The result of this research can be adopted into integrate query processing systems that integrate searching results on distributed digital objects and provide consistent views to users, without physical integration of digital libraries on the Internet. This system can be constructed by the conceptual integration of distributed and independent search systems of each digital library. Using the system will make easier to share and circulate information between special fields, areas or nations, etc.

## **2. Proposal of a framework to improve semantic interoperability**

We proposed a framework to solve semantic heterogeneity using three components. They are Metadata Registry (MDR), Schema Repository (SR), and Terminology Reference System (TRS).

## MDR

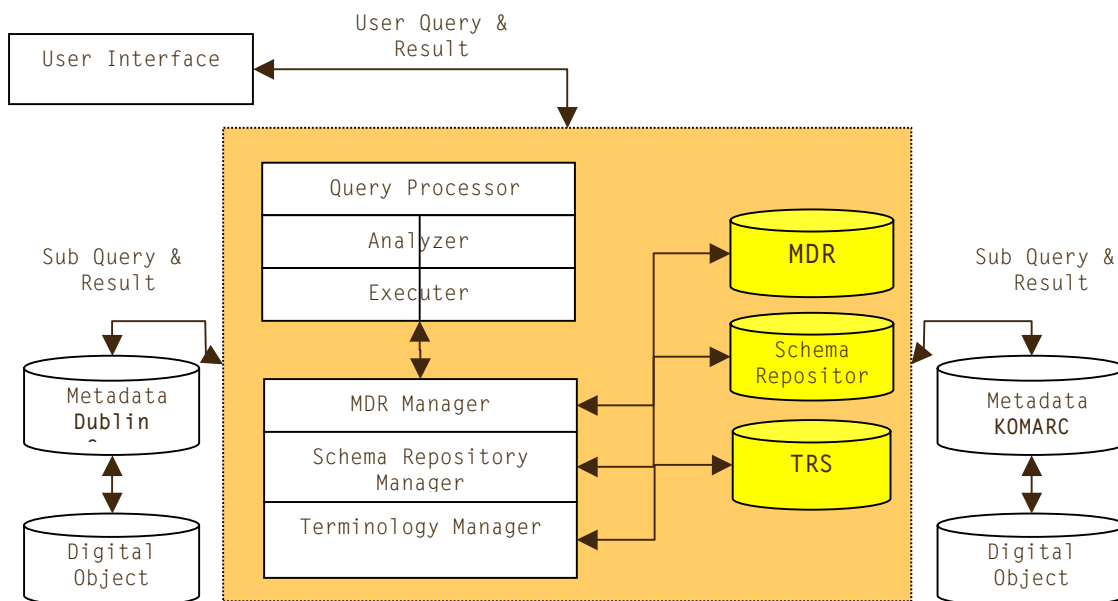
MDR is based on ISO/IEC 11179 : Metadata Registry. MDR is a system that registers and manages metadata. MDR plays a role of unique identification, registration and service of metadata. We intend to use MDR as a semantic coordinator of data element.

## SR

Schema repository, proposed in this research, contains schema information of local databases, such as table names, data element names, value domains, and referential integrities. SR maps the local schema to MDR according to its semantic. We can assume the attributes in local databases that have the same concept if they share the same data element concept in MDR. SR is the efficient solution to be able to show high correctness for end-user queries as well as to solve the semantic heterogeneity of data.

## TRS

TRS manages the heterogeneity of terms that are used in local databases and index databases. TRS also manages the relationships of terms, such as generalization, specialization, and inheritance. Queries are mapped to this TRS, thus enabling data to be integrated without requiring all resources to use the exact same terms.



**<Information retrieval system based on Metadata Registry>**