A Design of K-XMDR Search System Using Topic Maps

Zhang Jialei, Chi-Gon Hwang, Gye-Dong Jung, and Young-Keun Choi, Member, KIMICS

Abstract—This paper proposes a search system using the topic maps that it extends XMDR into Knowledge based XMDR for solving of the problems of the heterogeneity of distributed data on a network and integrate data by an efficient way. The proposed system combined Topic Maps and the extended metadata registry effectively. The Topic Maps represent related knowledge and reasoning relationship by associations of topic. And the extended metadata registry standards and manages the metadata of the local systems through registration and certification on the distributed environment. We also proposed a meta layer, include the meta topic and meta association to achieve semantic classification grouping of topics and to define relationship between Topic Maps and extended metadata registry.

Index Terms—Topic Maps, eXtended MetaData Registry, Knowledge-based System, Ontology, system integration.

I. INTRODUCTION

MOST information resources are distributed among many systems, and such information needs to be integrated in response to various demands. Therefore, managing distributed data in a heterogenic computing environment is the most fundamental and difficult.

Information managed by distributed DB system has the problem of semantic and structural heterogeneity, and therefore, sharing resources for inter-operation may cause many problems. When sharing information real time, in particular, the information sharing system should have a capacity to detect and resolve the collision of semantic or logical structure. In most cases, however, information sharing needs to modify the existing legacy system. Therefore, data integration to guarantee the independence of individual legacy systems is needed, and there is a need to manage semantic collision of information. As a way of resolving such collision, the concept of ontology has emerged. There are many different definitions of ontology by field that is used as an integral part of an intelligent system. Gruber defined ontology as “formal and explicit specifications for shared conceptualization of the field.”[1] As ontology has become a foundation of a knowledge-based system, the need of a language solely for ontology has been raised to correctly represent conceptualization structure, and standard languages such as RDF/RDFS[2], DAML[3], OWL[4], and Topic Maps[5] have been developed one by one. Among these, Topic Maps is a technical standard used to define the knowledge structure in a distributed environment and link this structure and knowledge resources, and can be said to be a new paradigm of the formation, extraction and navigation of information resources. XMDR is a system to integrate data by combining MDR and ontology to solve the problem of heterogeneity of distributed data and the technology to save XML-based relational DB meta-data in an object-oriented DB to resolve the heterogeneity of data as a result of data integration[6][7]. XMDR is, however, not enough to represent various kinds of ontology due to its lack of ontology representation and association. Therefore, we combine XMDR and Topic Maps as ontology, which is called TMDR, on the basis of which we propose a method of generating global query and converting it into local query.

In this study, Chapter 2 describes related literature review, Chapter 3 introduces representations of Topic maps and XMDR, Chapter 4 looks into how K-XMDR affects the entire system, and finally Chapter 5 is the conclusion of the study.

II. RELATED STUDIES

A. Topic maps

Topic Maps can connect one topic and another via association so as to have access to all knowledge information related to the theme represented by one topic.

Topic Maps can function as a link between knowledge representation and information management. Knowledge is clearly different from information; knowing something is different from having information about it. In this respect, knowledge management is reduced to three activities - ‘generation,’ ‘formalization,’ and ‘delivery.’ Topic Maps is the standard for formalization among them and essential to develop a tool for generation and delivery [5][10].

Components of Topic Maps are as follows [10]:

- Topic Class: also known as Topic Type, which means the classification of topics.

- Topic: It represents subject. Generating Topic that represent a certain theme in computer means that Topic convert the theme into an object which can be understood
and processed. Topic is instance of Topic Class.

Association: used to set the relationship between Topics defined in Topic Maps and thereby provide context of Topic in Topic Maps. Setting the relationship between Topics is essential to model knowledge.

Occurrence: a link of Topic-related information resources. When information resources provide information on Topic, it is represented as occurrence for the applicable Topic.


Like RDF/S and OWL, XTM 1.0 provides a way of realizing Semantic Web and is used to represent information of relational DB. While DB represents the relationship between information objects only, XTM 1.0 can connect different locations where information objects exist. However, most of the current tools to build Topic Maps are a stand-alone type that accesses an XTM and process it directly. This type can process XTM documents only, but it takes long to process XTM documents, and the type cannot process bulk Topic Maps data.

B. XMDR

For XMDR, standards should be set up by defining elements of DB of each node so as to prevent the heterogeneity of a system. Accordingly, XMDR secures the effectiveness of data by defining document structure. By making sure that change of schema in a node does not lead to random change of XMDR, credibility of data interchange should be secure.[8]

XMMDR is the result of combining MSO to prevent the heterogeneity of meta-data and ML with location info of each node and information on access authority as well as InSO to avoid the heterogeneity of actual data values [6].

Although XMDR is useful in resolving the problem of heterogeneity in schema among local data to integrate data, it has a limit in resolving the semantic heterogeneity by means of ontology. Therefore, as a way of representing ontology, K-XMDR combining Topic Maps that emphasizes association is used.

III. K-XMDR SYSTEM BY EXAMPLE

K-XMDR is a kind of Knowledge based systems which uses eXtended Metadata Registry to retrieve information in distributive network.

Knowledge based systems are artificial intelligent tools working in a narrow domain to provide intelligent decisions with justification. Knowledge is acquired and represented using various knowledge representation techniques rules, frames and scripts. The basic advantages offered by such system are documentation of knowledge, intelligent decision support, self learning, reasoning and explanation. Knowledge-based systems are systems based on the methods and techniques of Artificial Intelligence.

![Fig. 1. Structure of K-XMDR Search System.](image)

As in Figure 1, we divide the K-XMDR System into three layers: the User Interface, core of the K-XMDR Search System, and the Local System. The three layers of the framework can be connected and provide semantic interoperability with and integrate heterogeneous databases which are unable to be integrated in real time due to a number of problems arising from their heterogeneity. It provides user interface with easy access to a distributed system although without knowledge on heterogeneous data, by having the core of K-XMDR search system agent detect semantic collision of data in many distributed systems in real time and give a solution, and access the local system via a wrapper agent. Figure 1 depicts the overall system flow of this mechanism.

The system consists of 6 agents to carry out the operations of the system. The role of each agent is shown as below:

- Interface agent: Builds an interface for user access and provides a list of user-based knowledge search items.
- K-XMDR agent: processes K-XMDR and extracts not only the information needed, but also the information of changed query.
- DOP agent: receives the local query and local information which sent from Arbitration agent extracts the standard items at the global query.
- Arbitration agent: for related search queries, adjust all of the users’ needs and sends it to the interface agent.
- Wrapper agent: rewrites the query running on the local system for DB and sends the results to the gathering agent.
- Gathering agent: create a temporary table.

There are also 4 Database in this system. Proxy DB is to collect the results. And the role of other DBs will be described in later chapters.

A. Topic Maps of K-XMDR System

Topic Maps can solve the problem of data collision by defining the relationship of knowledge. There are a few
types of data collision: first, collision between actual data, and second, collision in schema to store data. Among them, Topic Maps is to deal with the second type of collision.

The association of topics of Topic Maps is as shown in Figure 2. For example, users enter search condition “Search the title and author of theses containing the key word ontology.” Topic Maps finds topics of which keyword is ontology, which can be included into search condition to raise the accuracy of the search. Figure 2 depicts the relationship of Topic Maps for this example.

As the component of Topic Maps appeared in the related study, there are topic class, topic, and association 3 layers in the Topic Maps of our proposed system. But in order to achieve the knowledge-based searching effectively, we propose a Meta layer, specific to the Topic Maps, is the Meta topic. The role of each layer are described as follow:

The Meta topic is a general classification of topics, the topic class is the specific classification, and the topic is the value of topic class. The association may only use between topic and topic, because it performs the relationship of topics. If we want to performance other relationships, such as relationship between Topic Maps and XMDR, we can use a Meta association. About the definition of meta association, we will explain it in the next section.

There are two reasons why we proposed the meta layer in an knowledge-based search system. First, we might suffer a lot of Heterogeneity situation at the distributed network environment. For example, there may be a paper named “XMDR”, and on the other hand, there also be another topic named “XMDR” which belong to the “interested area”. The proposed meta topic put forward a good solution to similar problems. Second, in an knowledge-based searching, users always hope to receive the reasoning result. We propose the meta topic to achieve this goal. For example, there is an existing topic call “Choi”, we have known that “Choi” belongs to the meta topic “Author”, so if there appears another topic which has some associations to “Choi”, we can reasoning the relationship between these two topics.

B. Meta Association: an association between Topic Maps and XMDR.

XMDR, which is to secure interoperability of data, manages meta-data standardized through meta-data registration and certification and maintains the compatibility between elements of meta-data by sharing specifications and meaning of meta-data. If different databases use different identifier or language for the same concept, an ontology that specifies the shared concept in a formal and explicit manner is combined to solve this problem. The ontology is to solve the collision between meta-data operated in many local systems.

C. Structure of global and local schema

In K-XMDR System, the main action between global schema and local schema is the standardization of the field name. We consider that the all the specific data is stored in the Local System. These data or information in different Local System have different expression. For example: different name, different content, or different range. So we have to set them standard. The specific solution will be shown in the next section.
Each Local System, the main direction of interest is also different as follow:

<table>
<thead>
<tr>
<th>Area</th>
<th>Schema structures and meta associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>Paper</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Local1</td>
<td>Paper, : association</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Local2</td>
<td>Paper, association</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Local3</td>
<td>Paper, : association</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Standardization process between global schema and local schema.

In figure 4, we use the standard file name “papername” as an example. At the first Local schema, we use the generic name “paper” to indicate that. And we can also express it as the acronym “pname” like Local 2, and the equal form “papername” at Local 3. In practice, we might encounter a more diverse expression at the local system. So we need to change them into a standard file name in the global schema. And the source the standard file name in our Topic Maps is the topic class. Therefore, we propose a Meta association so that the topic class and the standard file name can be connect and map effectively.

D. Data Structure of the K-XMDR System

In Figure 4, the ERD is a pivotal repository to perform the role of TMDR, which allows for the solution of the heterogeneous environment problem of the existing DB system.

The relationship of each table is described as follow:

1) Link between standard schema and meta association.
2) Link between topic class and meta association.
3) Link between meta topic and topic class (meta topic 1 ↔ topic class n).
4) Link between topic class and topic (topic class 1 ↔ topic n).
5) Association between topic and topic.
6) Role type of associations.
7) Association (include association members).

Fig. 5. ERD for K-XMDR.

E. Conversion of queries.

Global queries are similar to SQL except that they do not have a FROM clause. In SQL, standard items are used for fields. Standard items, which are global schema mentioned in Chapter 3 XMDR. A global query basically consists of a SELECT clause and a WHERE clause describing search conditions.

Converting global queries to local ones takes seven steps:
1) Create global queries.
2) Expand global queries by adding association via Topic Maps.
3) Parse global queries, extracting <gs_name> and <value> from global queries.
4) Create a temporary table to store the result in a Proxy. At this step, make a temporary table to collect results in Proxy using <gs_name> next to a SELECT sentence and session information of the user.
5) Extract TMDR information. This step is subdivided three many sub-stages, which are ① local schema extraction, ② extraction of synonym from associations of Topic Maps, and ③ extraction of local system access information of ML. <gs_name> searches information to convert into local items via XMDR. <value> in <condition> and <gs_name> become a pair and search a synonym in Topic Maps. To <value>, conditions can be added through semantic concepts and associations. In addition, by means of the relationship between global schema and local schema, access information of the local system which is managed by ML is extracted.
6) Add local item conversion and table info.

<standard field name> is changed into the extracted
local item. Here, a structural difference occurs, which can deal with combination and separation of strings. At this step, the table information of a local system is read using local items, and the From clause is included in a local query.

7). Apply association and semantic heterogeneity. At this stage, the content of Topic Maps extracted in Step 4 is applied to the local query, which expands the WHERE clause to increase the efficiency of the search.

**F. Data Flow of K-XMDR.**

In the previous section, we proposed the framework of the K-XMDR system. The framework can provide semantic interoperability with and integrate heterogeneous databases which are unable to be integrated in real time due to a number of problems arising from their heterogeneity. The problems may occur due to the difference of schema and that of representation of the same values such as data values, granularity, representation methods, modes, formats, and units. The framework provides users with easy access to a distributed system although without knowledge on heterogeneous data, by having the semantic agent detect semantic collision of data in many distributed systems in real time and give a solution, and access the local system via a wrapper agent. The figure of the framework depicts the overall system flow of this mechanism, and in this section, we will learn more about the data flow of our proposed system.

![Sequence Diagram](image)

**Fig. 6. The Sequence Diagram.**

Figure 6 is the sequence diagram of the K-XMDR system. This sequence diagram is generated on the basis of figure 1, which showed the framework of the system.

The process of data flow in the sequence diagram is as follow:

1. When a searching begin, the Interface agent sends a 1.standard interface create() request to K-XMDR agent.
2. The K-XMDR agent creates the Interface item and returns it to interface agent.
3. As we need to find the association in the Topic Maps, the interface creates a search() function to K-XMDR agent.
4. The K-XMDR agent finds out the association and returns it to interface agent.
5. Create a standard query in the Arbitration agent.
6. The created query is disassembled in the DQP agent.
7. The DQP agent creates a search () function to search the local information in K-XMDR agent.
8. After the local information return, the local queries are regenerated in the DQP agent.
9. Return the local queries into Arbitration agent.
10. Arbitration agent creates a result table in the proxy DB.
11. Local query is executed in the Local system (Wrapper agent).
12. After execution the local results return to the Gathering agent.
13. Gathering agent push the result to proxy DB.
14. Proxy DB create a result status notify to the Arbitration agent.
15. Result return to Interface agent.

**IV. APPLICATION AND SYSTEM COMPARATIVE ANALYSIS**

**A. The User Interface.**

In this chapter, we will test the results from implementation to test this study and experimental data. Figures 7 and 8 show how query is created by inputting the request of the user and how association is included into the query through Topic Maps.

![Main search interface](image)

**Fig. 7. The Main search interface.**
Figure 7 depicts the interface that helps users perform a search, which play functions in three domains; First, search item domain that selects required items from global schema, and thereby users can search even if they do not know about local system; Second, search condition domain where conditions are selected, that is, items and value of the condition. These two domains serve as data to create global query, which can be seen from the text on the right in Figure 7. These two domains serve as data to create a global query, which can be seen from the text on the right in Figure 7.

Figure 8 displays how association is searched on the basis of Topic Maps. Here, users provide the hierarchical structure of topics according to input item and value and the associations of topics selected from the hierarchical structure and thereby add the association to the query.

In this way, the K-XMDR system can have access to each local database without alteration.

![Topic Maps of K-XMDR System](image)

**Fig. 8.** Topic Maps that added association.

Figure 8 displays how association is searched on the basis of Topic Maps. Here, users provide the hierarchical structure of topics according to input item and value and the associations of topics selected from the hierarchical structure and thereby add the association to the query. In this way, K-XMDR can have access to DB in each local without alteration.

Following is a way of performance evaluation in which basic keyword search and association added search are compared.

The search process takes the following several steps:

Step1: Search Item Selection. In the first user interface, user can choose any topic class of each meta topic, then add the chosen topic class into the “Generate Global Query table”. (Can be multi-selection)

Step 2: Search Condition. In the Search Condition option, the chosen topic class would be standard set, in order to unify the different name of the various local systems.

Step 3: Select the Local System. There are several Local DBs that include the chosen topic class. Users can choose which Local DB they want to search. And then, put the “search” button to enter the topic selection interface.

Step 4: In the form of topic maps interface, users can select any topic that associate to the previously selected topic, and then, put the “apply” button to view the result.

Step 5: Query converting. The format of global query is similar to SQL sentence except that it has no FROM clause to designate the table to search. Field name uses global item of MDR. Global item is a virtual schema for schemata of the existing local systems.

Step 6: There might be a list which match with our selected keyword, and also match with the association of the keyword. For more importantly, the K-XMDR system can display the information of Co-author, and the reasoning result.

Step 7: The content of the final chosen topic would be displayed in the result interface. If there is more information on this item, it would be shown as expandable. If there is a topic in the content, it will be shown as underlined.

![Result of K-XMDR System](image)

**Fig. 9.** The first result interface(list).

![Result of K-XMDR System](image)

**Fig. 10.** The second result interface(detail).
After a searching, we can receive the result as a form of list which shown in figure 9. Users can get the information they want as well as the other information that have any relationship to them. And they can also get the reasoning result through the Topic Maps which shown at the bottom of figure 9.

If users interested in any information in the list above, they can also view the detail, which is shown as figure 10.

B. System Comparative Analysis

Current assessment of a particular product or small number of users in recommender systems based on further recommendations for a system using Knowledge-based search system has been studied. This system is compared with that from a survey of the system are as follows.

The proposed system uses XMDR to solve the problem that representation of heterogeneous data in a distributed environment. Thus, as shown in Table 1 above, the 6 items are compared to the system. The items are discussed in the Topic Maps support, XMDR support, storage addition on the structure and the type of save format, whether they are an knowledge-based search system, and whether they solved the heterogeneity and relationship problem.

| TABLE 1. SYSTEM COMPARATIVE ANALYSIS (C:Centralized, D:Distributed) |
|-----------------------|-----------------|--------|--------|
|                       | Key word Search | TM4L   | FAQ    | K-XMDR |
| Topic Maps support    | NO              | YES    | NO     | YES    |
| XMDR support          | NO              | NO     | YES    | YES    |
| Knowledge-based search | NO              | YES    | YES    | YES    |
| Heterogeneity and Relationship problem | Not solved | Relationshi p problem solved | Heterogenei ty problem solved | Both solved |
| Save format           | DB              | DB     | DB     | DB     |
| Storage Structure     | C               | C      | C      | C & D  |

About the 4 system in table 1, the Key word search is a familiar search system. And the TM4L and the FAQ system are both knowledge-based search system that appeared in the reference paper. Among that, the TM4L(Topic Maps for e-Leaning) [12] is an e-learning environment, which enables the creation, maintenance, and use of ontology-aware leaning repositories based on Topic Maps. And the FAQ (Frequently Asked Questions) [13] system Our overall conclusion is that FAQ pages on the web provide an excellent resource for addressing real users’ information needs in a highly focused manner.

By the system comparative analysis in table 1, we consider that, although the keyword search system is widely used by everyone, it is not suitable for knowledge search; it cannot solve the heterogeneity and relationship problem. On the other hand, the TM4L system has a topic maps support, so it can solve the relationship problem. And since the FAQ system supports XMDR to solve the heterogeneity problem effectively. However, our propose system could solve both relationship and heterogeneity problems, which often occurred in the distributed environment. The system proposed K-XMDR structure which use the Topic Maps and XMDR together effectively so that it provides a better way in a knowledge based search.

Although the save format of the four systems in table 1 are all database, the other 3 systems’ storage structure are centralized only, but our proposed system have their storage structure by the type of both centralized and the distributed.

Figure 11 shows the quantitative interpretation of these 4 systems. Through it, we can realize that the operation of keyword search is not complex, so it cost less time to complete a single search. But it has to interact with the uses at every time of search, so the keyword search has worse efficiency. On the other hand, the knowledge based search system has to spent more time to construct the knowledge database and to create the association between the elements, but it can provide a automotive way for searching, especially in many times of search.

IV. CONCLUSIONS

This study extends XMDR into K-XMDR (Knowledge-Based Extended Metadata Registry) to ensure the interoperability of data, and to describe multi agent system using K-XMDR. The proposal K-XMDR will be solved the problem of data heterogeneity and inconsistency in sharing and exchanging each local data in virtual pervasive computing environments. Data heterogeneity is generated by different definition or mismatched expression of the same meaning data. Therefore, we define and design K-XMDR to interoperate various each local data in pervasive computing. We also describe our synchronizing multi agent system, which uses K-XMDR to solve data heterogeneity for data interoperability through synchronizing data. It provides more efficient agent through reducing errors that
happened on synchronizing requests and adding new pervasive computing environments with their own each local data.

The realized system has advantages and effects as follows:

First, it can be applied as a way of converting queries in various data integrations, and it is possible to access local systems simply by converting queries under the existing system. Second, the user does not need to consider the metadata collusion when a business process is executed for cooperation between legacy systems. Third, the K-XMDR search system could search more useful information, not only the specific object, but also the related information of it.

Although this system has good performance in specific areas, we hope that it can achieve a greater range and more widely area. The proposed method is expected to improve by continuously applying to queries for further studies.

REFERENCES