Development of Contents Management System Based on Light-Weight Ontology

Kouji Kozaki, Yoshinobu Kitamura, and Riichiro Mizoguchi

***Abstract*—In the Structuring Nanotechnology Knowledge project, a material-independent platform for supporting development of innovative nano-materials is developed. The platform is called "Structured Knowledge Platform for Nano-materials and Products". It consists of various kinds of resources. However, domains and grain sizes of the resources are different from each other. Therefore, we need a consistent framework to interoperate them efficiently. In this background, we aim to develop a contents management system based on a light-weight ontology which provides a common vocabulary to represent semantic contents of the resources. This article outlines the system with its functionalities.**

***Index Terms*— ontology, content management, nanotechnology, semantic web**

I. Introduction

The research of nanotechnology is extended in various domains, and each domain closely intertwines with each other. Therefore, sharing the knowledge among different domains contributes to facilitate the research in each domain through cross fertilization. In this background, the Structuring Nanotechnology Knowledge project, which is a NEDO (Japanese New Energy and Industrial Technology Development Organization) funded national project [1], has been carried out. The goal of the project is to build a material-independent platform for supporting development of innovative nano-materials, which is called "Structured Knowledge Platform for Nano-materials and Products". It is neither a database, nor a set of simulation tools nor a knowledge base, but is an integrated environment composed of structured knowledge supported by advanced IT. The Structured Knowledge Platform consists of many resources such as conceptual design systems for "Process", "Structure" and "Function" of nano-materials, simulators, electronic textbooks, scientific papers/patents databases, and so on. These are developed as web-based systems.

Because these resources are developed by about 20 experts of nanotechnology and the development process are disjoint, the domain and the grain size of each resource are different from each other. Furthermore, a lot of resources are offered by other projects supported by NEDO and cooperative corporations. In addition to the resources, the number of documents and data in the databases is increasing day by day. The total of the resources amounts to a hundred thousand or more. Furthermore, each resource closely intertwines each other due to their complex semantic content.

To make full use of the platform, it is indispensable to manage the thousands of resources and interrelationships among them in a comprehensive way. Therefore, we need a consistent framework to manage and to make them interoperate efficiently. In this background, we aim to develop a web contents management system based on a light-weight ontology, which provides a common vocabulary to represent semantic contents of resources. This article outlines the content management system with its functionalities.

II. The Structured Knowledge Platform

A. The structure of the platform

The Structured Knowledge Platform consists of three layers: *data layer*, *information layer* and *knowledge layer* (Fig.1). The *data layer* contains structured databases of documents, patents, experimental data, and so on. These data in *the data layer* are analyzed and systematized in the *information layer* from several viewpoints. This layer consists of many resources such as various simulators, knowledge models, conceptual design systems for nano-materials, other knowledge systems, and so on. And the *knowledge layer* gives generalized knowledge and common bases for the platform.

In addition to the above resources, new resources have been developed as web-based systems and published on the website.

B. The position of the content management system

In the Structured Knowledge Platform, a comprehensive management system of nanotech knowledge has been developed to manage increasing resources (contents) and to promote efficient use of them.

Requirements for the system are *(1) to provide a standardized and efficient framework for management of*

Manuscript received December 31, 2006. This work was supported in part by the Japanese New Energy and Industrial Technology Development Organization under Grant the Structuring Nanotechnology Knowledge project, and the Japan Society for the Promotion of Science under Grant-in-Aid for Encouragement of Young Scientists (B) 17700150.

Kouji Kozaki is with the Institute of Scientific and Industrial Research (ISIR), Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan (phone:+81-6-6879-8416;fax:+81-6-6879-2120;e-mail:kozaki@ei.sanken.osaka-u.ac.jp).

Yoshinobu Kitamura is with the Institute of Scientific and Industrial Research (ISIR), Osaka University (e-mail: kita@ei.sanken.osaka-u.ac.jp).

Riichiro Mizoguhci is with the Institute of Scientific and Industrial Research (ISIR), Osaka University (e-mail: miz@ei.sanken.osaka-u.ac.jp).

resources (contents) in the platform, and *(2) to connect these resources each other systematically*.

The realization of them contributes to efficient extension of various resources and usability of the platform, and then it promotes new idea creation through cross fertilization.

On the basis of these requirements, following two systems have been developed.

1) A comprehensive management system of nanotech knowledge through index ontology

The system manages resources in the platform through Nanotech Index Ontology (NIO), which is a light-weight ontology to provide a systematized vocabulary of nanotechnology to express semantic contents of the resources. The semantic content of the resources is described as metadata with vocabularies defined in NIO. And the system adds the metadata to the resources for managing the resources through them. The system realizes a content-based retrieval using the metadata and NIO.

2) Knowledge Network system

The system represents and manages nanotech knowledge by network (graph) structure in a computer readable format. The representation of structured knowledge by a consistent format can manage the knowledge of various domains consistently. And then, the system can connect resources of respective domains through following links in the network. In the system, nodes represent fundamental knowledge and links represent relationship of them. The fundamental knowledge is managed using NIO.

Common ideas among these two systems are to use computer readable format for representing semantic contents of resources and to unify the ways of contents management through them. In the following sections, the authors describe a prototype system of the comprehensive management system of nanotech knowledge through index ontology.

III. CONTENTS MANAGEMENT TRHOUGH LIGHT-WEIGHT ONTOLOGY

In this section, we describe the index ontology and the metadata to manage the resources. And then, we present a frame work for contents management through the index ontology.

A. *Index ontology*

If semantic contents of the resources are described in free keywords or a natural language, it is difficult to manage them by a computer system comprehensively because of their diversity. We avoid such a problem using common vocabulary to represent the semantic contents of resources in metadata. The common vocabulary is defined in a light-weight ontology, which is called *Index Ontology*.

The index ontology is an ontology which is specialized to use as “index” for contents management, and it is represented systematically by concepts (vocabularies) and *is-a* (generic-specific) relation among them. We suppose to represent the semantic contents of the resources in enumeration of the concepts which is defined in the index ontology so that we represent them simply to process them with the system easily and efficiently. In this purpose, we consider the index ontology is enough to represent the semantic contents.

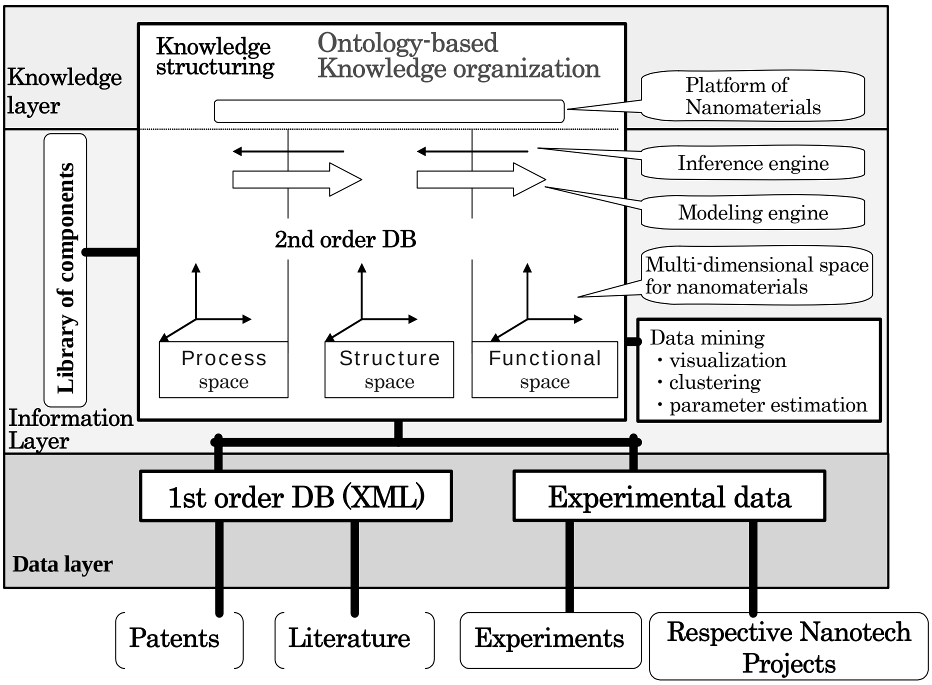

Fig.1. The Structured Knowledge Platform [1]

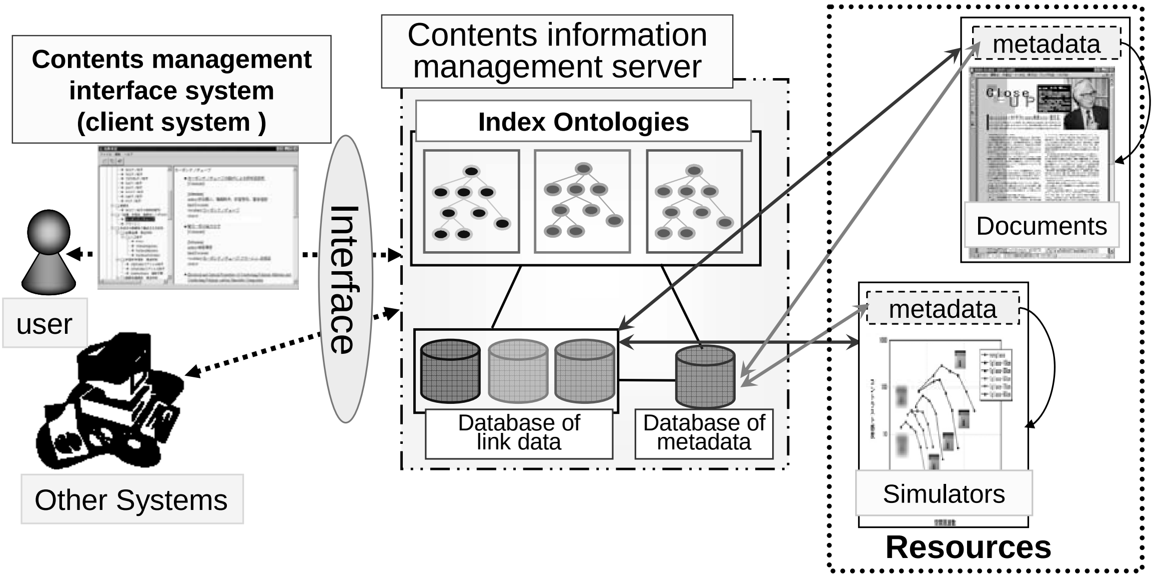

Fig.2 The architecture of contents management system

B. Metadata to represent semantic contents of resources

Here, we consider metadata to represent semantic contents of the resources for contents management. As the most fundamental requirements for a contents management system, two functionalities are required. One is to classify the resources by semantic content of them and make it possible to access the resources easily through the classification. The other is to connect related resources each other based on their semantic content.

For this purpose, we introduce two kinds of metadata: *main topic* and *sub topic*. The metadata of main topic describes the subject of the resources with a few keywords. And the metadata of sub topic describes related topic of the resources with keywords. These keywords are selected from the concepts defined in the index ontology. We call the concepts selected as keywords in metadata of main topic *"subject concepts"* and the concepts selected as keywords in metadata of sub topic *"related concepts"*. The subject concepts are used to classify the resources and access to the resources, and the related concepts are used to connect related resources.

For example, the subject concepts of a resource about "synthesis of carbon nanotube" are "carbon nanotube" and "synthesis". And then, we suppose that its related concepts are "strength", "electrical property", and so on because they are features of carbon nanotube.

C. Contents management through light-weight ontology

We developed a framework for contents management through the index ontology and the metadata discussed previous section. In our framework, every resource is added the metadata of main topic, sub topic, and so on. The framework realizes two functionalities (1) to manage resources based on its subject concepts, and (2) to connect related resources through its related concepts.

1) The framework to manage resources based on its subject concepts

The content management system classifies the resources by subject concepts in its metadata of main topic, and then the system gets a list of resources whose main topic includes same concept as its subject concepts for every concept in the index ontology. We call the list of resources *"link data"* and use it to manage the resources.

Every concept in the index ontology has a link data[1], and it keeps a list of resources which includes the concept in main topic. When a system needs information of resources which include a concept as main topic, the system accesses to link data of the concept and gets the lists of such resources.

2) Framework to connect related resources through its related concepts

It is important to manage relations among resources in the Structured Knowledge Platform because the research of nanotechnology is extended in various domains and each domain intertwines with each other closely. In our framework, the relations among resources are managed through lists of related contents in its metadata of sub topic.

When a system needs information of concepts related to a resource, the system accesses to its metadata of sub topics and gets the lists of related concepts. And then the system can connect related resources through the link data corresponding to the concepts.

IV. Development of Contents Management System

A. The architecture of the system

We developed a prototype of the contents management

[1] Some concepts don't have resources which select the concept as subject concept. The link data of such concept is treated as a null set.

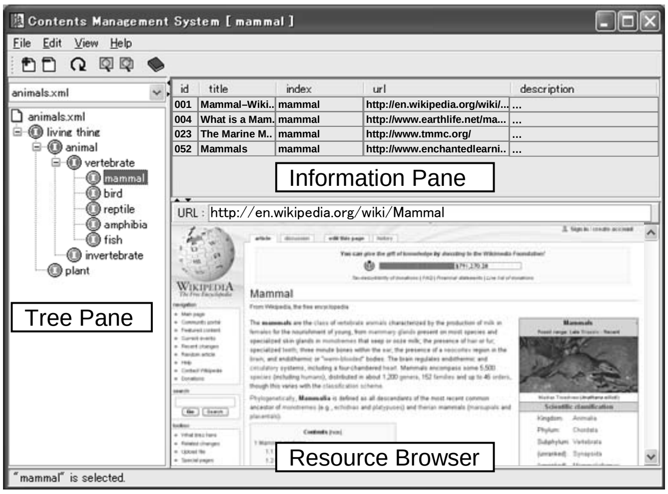

Fig.3 The contents management interface system

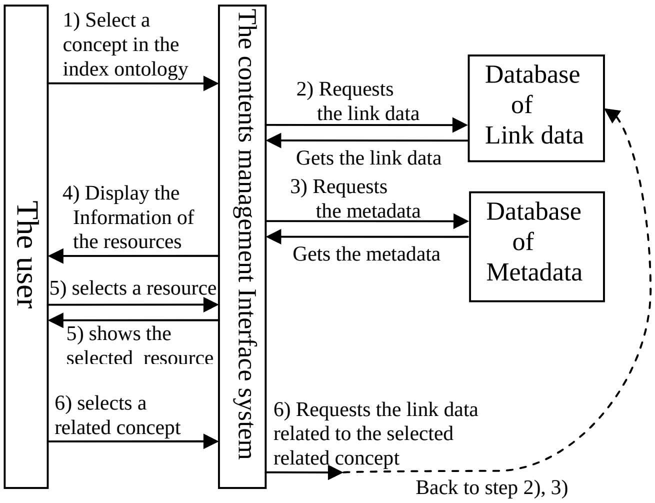

Fig.4 How to work the contents management interface system

system based on the framework discussed in section III. Fig.2 shows the architecture of the system. It is composed of *contents information management server* and *contents management interface system*. The contents information management server manages resources through index ontology by using link data and metadata. It is connected other systems through an interface, and it provides information of resources with requests. The contents management interface system is a client system to browse and modify the link data and the other metadata with graphical user interface. The users can access information in the contents information server with the system.

These systems have been developed in Java, and protocol to connect the server and the client is http.

B. Contents information management server

The contents information management server consists of index ontology, database of link data, and database of other metadata.

1) Index ontology

We used Hozo[2] [2, 3], which is ontology editor developed by the authors, to develop the index ontology and stored it in the server system. The system access it thorough HozoAPI [4] which is an API to use ontlogies built by Hozo. When the system uses an ontology in RDFS[3] or OWL[4] format as an index ontology, it can access it through Jena [5] instead of HozoAPI.

2) The database of link data

Each concept in the index ontology has one link data. The link data are stored in a SQL database. The link data consists of information about resources which includes the concept as subject concept. The information includes the title of the resource, its URL, and brief description of it. The URL is used as database key to identify resources.

3) The database of metadata

Metadata of resources is not added directly to each resource but stored with reference information of the resources in the database of metadata. The database is also developed in SQL. Because the system manages all metadata independently of resources, the system can treat not only metadata of resources in the platform but also that of outside resources such as other web pages. The metadata includes a list of related concepts, its authors, a kind of resource, and its source.

Interfaces to access these two databases are developed as web services implemented in Java servlets. Other systems can get data in the databases by using commands for requests to the services through http protocol.

C. The contents management interface system

Fig. 3 shows a snapshot of the contents management interface system. The system display information of resources in the contents information management server with a request of the user. The user can add, delete and modify the information of resources, link data and metadata in the server with the interface system. The system is composed of *tree pane*, *information pane*, and *resource browser*. The tree pane shows the index ontologies in a hierarchy structure. The user can select the concept of the index ontology in the hierarchy, and then the system get a list of subject concepts form the database of link data in the server through servlets and show the lists in the information pane. At the same time, the system gets and displays the information of metadata of each resource in the list. The user can access to resources displayed in the information pane by clicking the title of resource. The resource browser shows the selected resources.

The authors describe the outline of how to work the system in the following (Fig.4):

1) The user selects a concept in the index ontology

[2] http://www.hozo.jp
[3] http://www.w3.org/TR/rdf-schema/
[4] http://www.w3.org/TR/owl-features/

2) The system requests and gets the link data of the concept from the database of link data.
3) The system gets a list of subject concept from the link data. And then, it requests and gets the metadata related to each resource from the database of metadata.
4) The system displays the information of resources which are obtained in step 2) and 3) in the information pane.
5) If the user selects a resource displayed in the pane, the system shows the resource in the resource browser.
6) If the user selects a *related concept* in the pane, it displays the information of resources related to the concept obtained through its link data by the same steps of 2) to 4). And then, the user can learn related resources to see the displayed information.

D. *The utilization in the Structured Knowledge Platform*

In this section, the authors describe the utilization of the contents management system in the Structured Knowledge Platform. In the Structuring Nanotechnology Knowledge project, we built a preliminary ontology of nanotechnology in collaboration with domain experts based on keywords which are extracted from textbooks, papers and patents. We call the ontology "Nanotech Index Ontology (NIO)". It is constructed from the name of concepts in an *is-a* hierarchy. These concepts are categorized into 5 categories (Process, Structure, Function, Material, and Application) and the number of them is about 2,300. We used the ontology as index ontology for the contents management system and developed the system as an implementation of the system discussed this section.

Fig.5 shows the contents management interface system in the project. We call the system *NIO management system*. It was developed as extension of the contents management interface system in the some functionality.

The extended functionalities are:

- A functionality to switch 5 categories of NIO in the display.
- Overlooking nanotech knowledge in the platform.
- A functionality to display the information of resources related to a concept in NIO with a query from other system in the platform.

In the current version of the system, the user can overlook scopes of *the Nanoparticle project* and *the Nanocarbon project* which are collaborative project of our project. The system supplies the users a portal to the platform.

Furthermore, the concepts in NIO are used as common vocabulary by other knowledge systems such as Knowledge Network System and a Creative Design Support system of Nanomaterials [6]. By cooperating with these systems, the contents management system could realize more advanced management of resources.

V. Future Work

A. *Applications of the system*

Because a target domain of the content management system depends only on a domain of the index ontology, the system can manage various contents by using an appropriate ontology. And the framework is very simple and applicable to not only web contents, but also other kinds of contents. Therefore, the system is all-purpose and usable for a large variety of applications to manage various kinds of contents such as: publications, achievements, reports, references, pictures, bookmarks of website, and so on. To manage them more efficiently, some extension of the system is needed according to a purpose.

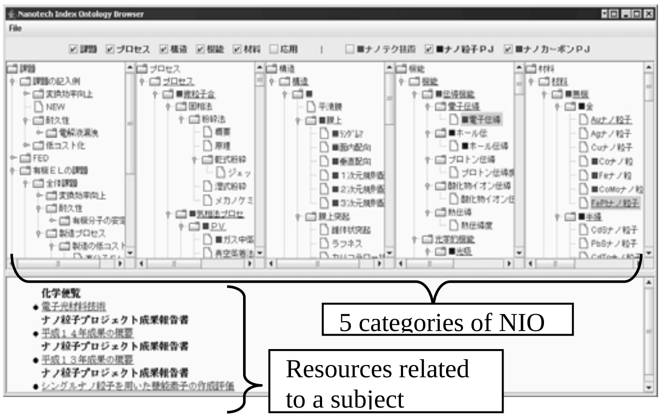

Fig.5 NIO management system

The system discussed in section IV is an example of the application to manage results of a project about nanotechnology.

B. *A large scale application*

The authors developed a prototype of the contents management system and tried to use the system in a small scale example. And now, we plan to use it in a large scale application. The Structured Knowledge Platform has more than 100 thousands resources in the server system. We have tried to use the system to manage a part of them. And then we are developing a contents management system for the whole of them.

C. *An improvement of index ontology*

By using a more detailed ontology as index ontology, it is possible to manage resources based on deeply understand of its semantic contents. A detailed ontology consists not only *is-a* relations but also *whole-part* relations, attributes of concepts, axioms and so on. For example, if attributes of concepts are defined in an index ontology, the system can find resources which have attributes specified by users.

In the Structuring Nanotechnology Knowledge project, we plan to improve Nanotech Index Ontology and to develop a frame work for contents management thorough them.

D. *An improvement of metadata*

There are standard formats of metadata such as Dublin Core [7]. But we consider semantic contents of resources are more important than format of information. Therefore, we plan to structured knowledge based on semantic contents as metadata.

For example, we think Knowledge Network (see section II) is usable as metadata for content management.

E. A management of viewpoint

By managing link data and metadata of each user separately, the system can manage contents according to a viewpoint of each user. Furthermore, it can manage contents based on an integrated viewpoint by unifying the databases of the users. If the index ontology is shared by all users, the integration is easy to realize. But if the users use different ontologies, a technique for ontology integration is needed.

F. Coordination with web crawling tool

In our system, some users have to input the information about contents into the system and it is concentrated in the contents information management server. Therefore, it takes costs to store the information in the system, and a mechanism to store the information more efficiently is needed. We plant to develop a mechanism that a web crawling tool collects information about contents and stores it in the server automatically. It means that we apply our framework to web resources which has some semantic metadata for Semantic Web.

VI. CONCLUSION

The authors discussed a contents management system based on light-weight ontology and its utilization in the Structuring Nanotechnology Knowledge project. The system manages various resources comprehensively thorough index ontologies and metadata which provide systematized knowledge to represent semantic content of the resources. It realizes a contents management from a unified viewpoint and contributes to use of the resources efficiently across various domains. We plan to use this system in an enterprise and extend the system for contents management with some intelligence.

ACKNOWLEDGMENT

We are grateful to the members of the Structuring Nanotechnology Knowledge project for their valuable discussions. And we are grateful to Mr. M. Ohta for their support to implement our system.

REFERENCES

[1] Structured Knowledge Platform for Nano-materials and Products, http://mandala.t.u-tokyo.ac.jp/

[2] Kouji Kozaki, Yoshinobu Kitamura, Mitsuru Ikeda, and Riichiro, Mizoguchi, Hozo: An Environment for Building/Using Ontologies Based on a Fundamental Consideration of "Role" and "Relationship",Proc. of the 13th International Conference Knowledge Engineering and Knowledge Management (EKAW2002), pp.213-218, Spain, October 1-4, 2002

[3] Kouji Kozaki, Yoshinobu Kitamura, Mitsuru Ikeda, and Riichiro Mizoguchi, Development of an Environment for Building Ontologies Which Is Based on a Fundamental Consideration of "Relationship" and "Role", Proc. of the Sixth Pacific Knowledge Acquisition Workshop (PKAW2000), pp.205-221, Sydney, Australia, December 11-13, 2000

[4] Kouji Kozaki, Yoshinobu Kitamura and Riichiro Mizoguchi, Developing Ontology-based Applications using Hozo, The Fourth IASTED International Conference on Computational Intelligence (CI2005), Calgary, Canada, July 4-6, 2005

[5] Jena – A Semantic Web Framework for Java, http://jena.sourceforge.net/

[6] Kouji Kozaki, Yoshinobu Kitamura, and Riichiro, Mizoguch, Systematization of Nanotechnology Knowledge Through Ontology Engineering－A Trial Development of Idea Creation Support System for Materials Design based on Functional Ontology, Poster Notes of the Second International SemanticWeb Conference (ISWC2003), Sanibel Island, FL, USA, Oct. 2003

[7] DCMI：Dublin Core Metadata Element Set, Version 1.1: Reference Description <http://dublincore.org/documents/dces/> , 1999, DCMI Recommendation