MATERIAL DATABASE SYNDICATION WITH RSS

Toshihiro Ashino[1]* and Nobuto Oka[2]

*[*1]Dept of Regional Development Studies, TOYO University, 1-1-1 Izumino, Itakura-machi, Ohra-gun, Gunma 374-0193 Japan*

Email: ashino@acm.org

[2]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583 Japan

Email: nobuto.oka@gakushikai.jp

ABSTRACT

Every material database on the Internet has a different data schema. There are some trial attempts to define unified schema for material databases, but since the structure of scientific data is very complicated and changes dynamically, defining a complete data schema is an impossible task. There are two major approaches for material data standards: one is MatDB, an attempt to define precise and detailed metadata, and the other is MatML, which only defines a framework. However, there is a third way: loose syndication, such as blogs with RSS. RSS, RDF Site Summary, or Really Simple Syndication, was developed to summarize document pages, but it can be extended to describe metadata of factual databases. In this presentation, an RSS extension for material database summaries is discussed.

Keywords: Material database, Metadata, RSS, Markup Language, MatML, MatDB

1 INTRODUCTION

Many material databases concerned with a variety of properties and materials have been developed independently. It is fundamental to promote the practical application of material databases in which the core portal of material data is built up, and data in a universal format are supplied continuously. For this purpose, standard material data formats, e.g., MatML [MatML] or MatDB [NMC-MatDB] have been developed. In the field of life science, there is a system for cooperation in a wide range of fields from genomics to protein information. However, in the field of material science, such an effort is hardly adequate.

Concretely speaking, a universal format is standardized by means of inheriting data formats of existing databases. It needs to expressly provide necessary fundamental information in a compatible format. Furthermore, the structure should be able to make adjustments as needs arise because the data structure for science and technology is overly complicated and dynamically transforms every time something new is discovered.

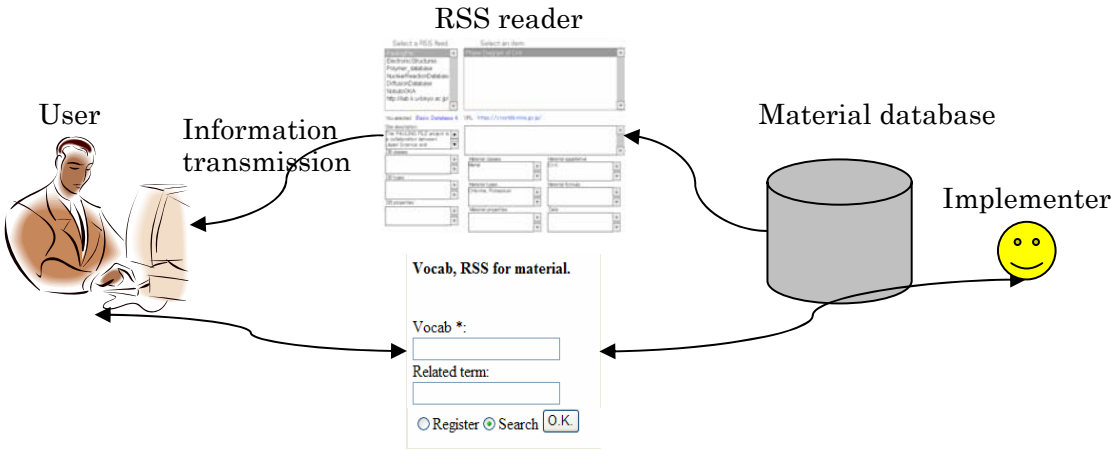
In order to promote effective utilization of existing material databases, an information transmission portal and a

standard for metadata description of material database should be established.

2 DEVELOPMENT AND OPERATION OF AN INFORMATION TRANSMISSION PORTAL FOR MATERIAL DATABASES

We regard "RSS," which has become widespread in information transmission of blogs and news sites, as an important technology when a portal for material databases is developed. Though there are currently several versions of RSS, RSS (RDF Site Summary) 1.0 (Beged-Dov, et al., 2001) was adopted in this research. The reasons include that RSS 1.0 engages for the core vocabularies and the extensibility by defining commonly-used vocabularies and elements as "core" and the others as "module." In addition, it is based on RDF, so the data structure is easy to change as needs arise. There are several trials to extend RSS distribution of metadata of scientific data in geoscience (Kubo, et al., 2006) and chemical science (Murray-Rust, et al., 2004; Rzepa, et al., 2006). However, there is no approach to extend RSS for material databases.

A RSS module is proposed for material information, and a portal is developed for transmitting information of material databases via networks. The portal consists of an "RSS reader" and "Registration function of available vocabularies in RSS module for material information" (Figure 1). An RSS reader is application software for obtaining RSS feeds. The implementer of each material database creates RSS feed, which describes its database following the RSS extension. The RSS reader aggregates these RSS feeds from database sites. RSS does not, however, define vocabularies of material properties or types, which are used to search and identify material databases. They should be defined in addition to RSS tags.



Registration function of available vocabularies in RSS module for material information

Figure 1. System construction of information transmission portal for material databases

2.1 RSS module for material information

Material names and classifications are denominated under the object domain, specification, nomenclature system (UNS (ASTM, 2004), IUPAC (Leigh, et al., 1998)), constituent element, crystal structure, binding state,

microstructure, shape, use application, property, functional capability, raw material, process of manufacture (location, manufacturer), sociality, developmental project, etc. There may be a variety of classifications depending on circumstances, but the variety makes cooperation between databases difficult. Consequently, an RSS module for material information defines only requisite elements when a user of the portal searches the databases' metadata; *class*, *type*, *name*, *composition*, and *property* were extracted as important tags as shown in (Table 1). They are selected as fundamental concepts to find out a material database.

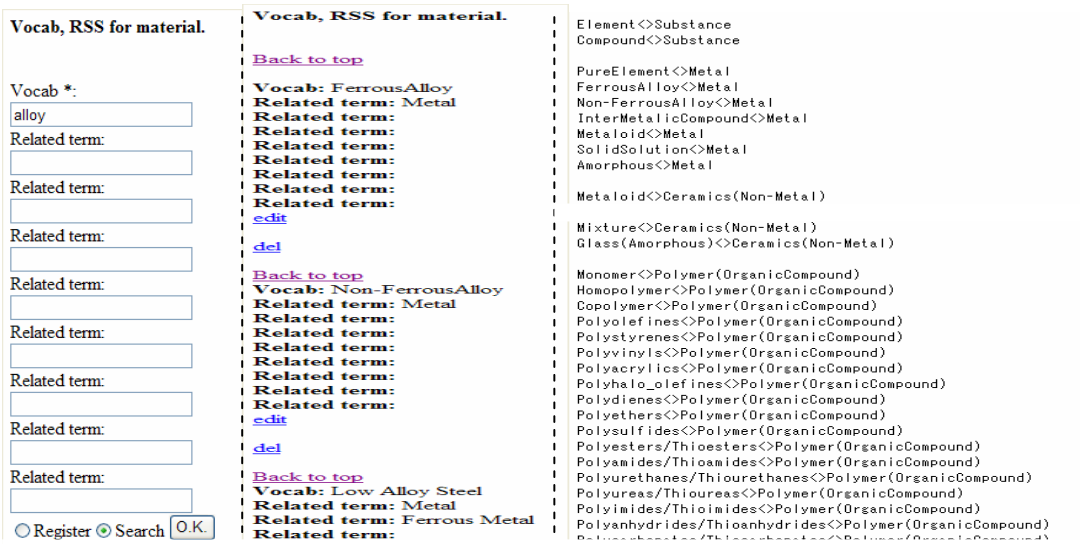
On the other hand, diversification of vocabularies in *class*, *type*, and *property*, which describe material information, becomes problematic at the moment of an information search. While many solutions, such as an expression method making use of thesaurus, are proposed, a unique undisputed method does not exist. Hence, it is necessary to confine available vocabularies and prepare a framework for fulfilling wide-ranging requirements.

Table 1. An example of RSS module for material information; "ma" is a XML namespace of this RSS module.

RSS tags	Definition	Examples
ma:class	Classifications based on material structures.	Ceramics, Polymer
ma:type	The other classifications.	Ferrous alloy, Carbon Steel
ma:name	Material names.	Mg-O, SUS304
ma:composition	Material composition.	Mg, O, Fe, C
ma:property	Properties included in material databases.	Crystal Structure, Phase Transition

2.2 Available vocabularies in RSS module for material information

Available vocabularies were selected from Matdata.net [MatData.net] and MatNavi of NIMS [MatNavi] as standardized vocabularies of the first version of the RSS module. Adopting generally known vocabularies makes it possible to enhance more data sharability and interoperability. As a result, vocabularies of *class* are Substance, Metal, Ceramics (Non-Metal), Polymer (Organic Compound), Composite, Natural, and Nanomaterial. In contrast, *type* and *property* need to fulfill various requirements of database implementers and users. For this reason, a web-interface was developed, where they can register and then search available vocabularies and related terms of *type and property* via the Internet (Figure 2). This is a framework for using a broad range of relationships among available vocabularies and related terms in order to improve user convenience of information searches. Furthermore they enable the creation of a database map.



(a) (b) (c)

Figure 2. Registration / search function of available vocabularies in RSS module for material information (a) Registration / search function of available vocabularies, (b) a result of the vocabulary search, (c) list of registered vocabulary

2.3 RSS reader for RSS module of material information

Currently in Japan, we transmit RSS feeds about sites' information for 151 existing material databases (Figure 3), drawn up on the basis of this suggested method, and develop an RSS reader for obtaining these RSS feeds (Figure 4).

```
<channel rdf:about="http://crystdb.nims.go.jp/">
    <title>Basic Database for Crystal Structures _ Pauling File</title>
    <link>http://crystdb.nims.go.jp/</link>
    <description>The PAULING FILE project is a collaboration between Japan Science and
    Technology Corporation (JST) and Material Phases Data System (MPDS)...</description>
    <ma:class>Composite</ma:class>
    <ma:type></ma:type>
    <ma:name></ma:name>
    <ma:composition></ma:composition>
    <ma:property>Crystal Structure</ma:property>
    <items>
        <rdf:Seq>
        <rdf:li rdf:resource="https://crystdb.nims.go.jp/" />
        </rdf:Seq>
    </items>
</channel>
```

Figure 3. An example of RSS feeds for a material database.

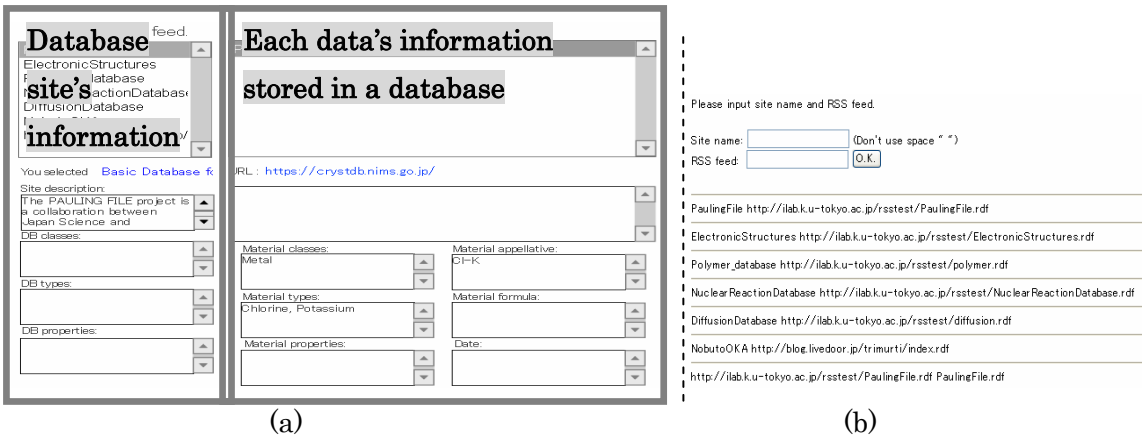


Figure 4. RSS reader; (a) RSS reader for material information and (b) registration function of RSS feeds.

This RSS reader has two windows; one shows the database site's information, and the other shows each data's information stored in a database. Users can obtain information on various material databases through it.

3 CONCLUSION

An RSS module for material information, which has been proposed in this paper, makes it easy to set up a material database portal. Standardized vocabularies of material classifications and properties enable the creation of a map of material databases with specified keywords. Currently in Japan, we transmit RSS feeds about sites' information for 151 existing material databases, drawn up on the basis of this suggested method.

4 REFERENCES

ASTM (Corporate Author) (2004) Metals *& Alloys in the Unified Numbering System (UNS), 10th Edition*. Warrendale, PA: SAE International.

Beged-Dov, G., et al. (2001) RDF Site Summary (RSS) 1.0. Retrieved November 15, 2006 from the World Wide Web: http://web.resource.org/rss/1.0/spec.

Kubo, T., Murata, K., Kimura, E., Ishikura, S., & Shinohara, I. (2006) *Automatic Meta Data Collection System for Satellite and Ground-based Observation Data by the STARS RSS1.0: An approach for the STARS Semantic Web*. CODATA2006, Beijing.

Leigh, G., Favre, H., & Metanomski, W. (1998) *Principles of Chemical Nomenclature: A Guide to Iupac Recommendations (IUPAC Chemical Data Series)*. Oxford: Blackwell Science.

MatData.net. Retrieved November 15, 2006 from the World Wide Web: http://www.matdata.net/

MatML. Retrieved November 15, 2006 from the World Wide Web: http://www.matml.org/

MatNavi. Retrieved November 15, 2006 from the World Wide Web: http://mits.nims.go.jp/matnavi/

Murray-Rust, P., Rzepa, H., Williamson, M., &. Willighagen, E. (2004) Chemical Markup, XML, and the World Wide Web. 5. Applications of Chemical Metadata in RSS Aggregators. *Journal of Chemical Information and Computer Sciences 44 (2)*, pp. 462-469.

NMC-MatDB. Retrieved November 15, 2006 from the World Wide Web: http://www.nims.go.jp/vamas_twa10/AMM_DB/XML-Schema.LZH

Rzepa, H., Wheat, A., & Williamson, M. (2006) ChemSem: An Extensible and Scalable RSS-Based Seminar Alerting System for Scientific Collaboration. *Journal of Chemical Information and Modeling 46 (3)*: 985-990.