

DEVELOPMENT OF AN OPENMATH CONTENT DICTIONARY FOR MATHEMATICAL KNOWLEDGE OF MATERIALS SCIENCE AND ENGINEERING

Toshihiro Ashino^{1*} and Yuichiro Yamashita²

¹*Toyo University, 2-36-5 Hakusan, Bunkyo-ku, Tokyo 112-0001, Japan*

Email: ashino@acm.org

²*National Institute of Advanced Industrial Science and Technology in Japan, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*

Email: yuichiro-yamashita@aist.go.jp

ABSTRACT

Many relationships between parameters and physical properties in materials science and engineering are represented as mathematical expressions, such as empirical equations and regression expressions. Some materials databases handle such information with indirect methods: as a table of sets of parameters, as a list of statements of programming languages, and other ways. There is no standardized way to represent mathematical relationships, and that makes it difficult to exchange, process, and display such information.

The AIST (National Institute of Advanced Industrial Science and Technology in Japan) thermophysical property database manages sets of parameter values for expressions and Fortran statements that represent relationships between physical parameters, e.g., temperature, pressure, etc. and thermophysical properties. However, in this method, it is not easy to add new parameters, to process expressions, and exchange information with other software tools. In this paper, we describe the current implementation of representing mathematical knowledge in the AIST thermophysical property database, and we also discuss its problems, sample implementations, and definitions of the OpenMath content dictionary for materials science and engineering.

Keywords: Mathematical Knowledge, Materials Database, OpenMath, Expressions

1 INTRODUCTION

Mathematical expressions are used to describe relationships between parameters and are important in the estimation of physical properties and behaviors of materials (Fukuyama & Waseda, 2009; Yamaguchi & Itagaki, 2002; Jasper, 1972). Some materials databases attempt to manage and provide such information (Baba, Yamashita, & Nagashima, 2009), but there is no standardized way to represent this information in the areas of materials science and engineering. Materials databases all use their own formats. There are two aspects which should be considered. One is presentation, how to render a mathematical expression on a screen. The other is semantics, which is required for processing, e.g., drawing figures or calculating values of expressions. In many cases, because materials databases are developed using relational database management systems (RDBMS), mathematical expressions are managed as tables or in text form.

There are two major methods of handling expressions in RDBMS that correspond to these two aspects. One is to manage a list of parameters for pre-defined mathematical expression types. The other is to manage statements of Fortran or other programming languages as text data. The former method is suitable for rendering mathematical expressions on Web pages or converting them into other formats, such as Fortran statements, because the correspondence of a parameter and a member of an equation representing the semantics of an expression is clear. This method is less extensible because in order to add new expression types, the structure of the database must be modified. On the other hand, the latter method can handle any type of expression that can be written as a statement of a programming language and can easily be used to calculate values by using the text data as a part of the program. It is difficult, however, to extract the semantics of an expression, such as the names of parameters, structures of

expressions, and restriction conditions. This requires parsing the programming language and analyzing the syntax tree.

The AIST thermophysical property database employs both methods to handle mathematical relationships between thermophysical properties and physical parameters. Figure 1 shows the data entry screen of the AIST database for a mathematical equation. Users select from pre-defined expression formats, Type1 or Type2, and define a parameter set. Parameter sets are managed as tables of the relational database. This means that when a new expression type is added, it must be built into the system. Adding new parameters may affect the database schema. In order to allow users to register free format equations, the system provides a function to manage equations as Fortran or Visual Basic statements. These two data formats, however, are not linked each other.

j	Variable	ISO	Unit	Minimum	Maximum	α_j	β_j
1				0	0	0	0

i	A_i	γ_{ij}
1		0

Figure 1. Equation data entry screen of the AIST thermophysical property database, which allows selection of a pre-defined equation formats and parameter setting

2 THE OPENMATH STANDARD

OpenMath is a standard for representing the semantics of mathematical objects that is developed and maintained by the community of mathematics (Buswell, Capriotti, Carlisle, Dewar, Gaetano, & Kohlhase, 2004). Because it is designed to exchange mathematical objects between heterogeneous mathematical systems, OpenMath descriptions can be handled easily by computer programs. On the other hand, another standard, MathML (Carlisle, Ion, & Miner, 2010), a standard developed by the Web community, mainly represents presentations of mathematical objects. MathML also defines its own markup syntax for mathematical meaning, called content markup. The content markup of MathML version 3 is intended to be compatible with OpenMath.

In OpenMath, common definitions of mathematical symbols are stored in a content dictionary (CD). The standard defines a notational method using XML, and in this case, each CD has its own namespace, and each symbol has its own URI. This means symbols defined in CDs can be referred to from other materials data formats in XML, such as Materials Ontology (Ashino, 2010) and MatDB (Ojala & Over, 2008). Also, mathematical objects, such as variables, parameters, and expressions, are able to have attributes such as units and types. Many CDs have been developed and provided by the OpenMath Society, including definitions of fundamental physical constants, dimensions, and units, which can be the basis for describing, processing, and exchanging materials knowledge. Many software tools, mathematical systems, and converters between OpenMath, MathML, and LaTeX are available.

3 CONTENT DICTIONARY FOR THE AIST THERMOPHYSICAL PROPERTY DATABASE

The OpenMath XML notation for the AIST database is divided into two parts. One is a CD named “aist_tpdb1” that defines the generic format of equations, parameters, variables, and mathematical operators, each annotated with its units. The other part is a RDF file, which references the format of equations defined in “aist_tpdb1” by their URI and specifies the values of the parameters that are derived from experiments and statistical data analysis. This means each equation defined in “aist_tpdb1” can have multiple sets of parameter values, which are calculated from different data sets.

Figure 2 displays an OpenMath entry for Eq. (1), which estimates the emissivity ε of liquid UO_2 (Fink, 2000).

$$\varepsilon(\lambda = 630\text{nm}) = 1 - 0.16096 \exp[-3.7897 \times 10^{-4} \Delta T - 3.2718 \times 10^{-7} (\Delta T)^2] \quad (1)$$

for 3120K ≤ T ≤ 4200K and 400nm ≤ λ ≤ 700nm

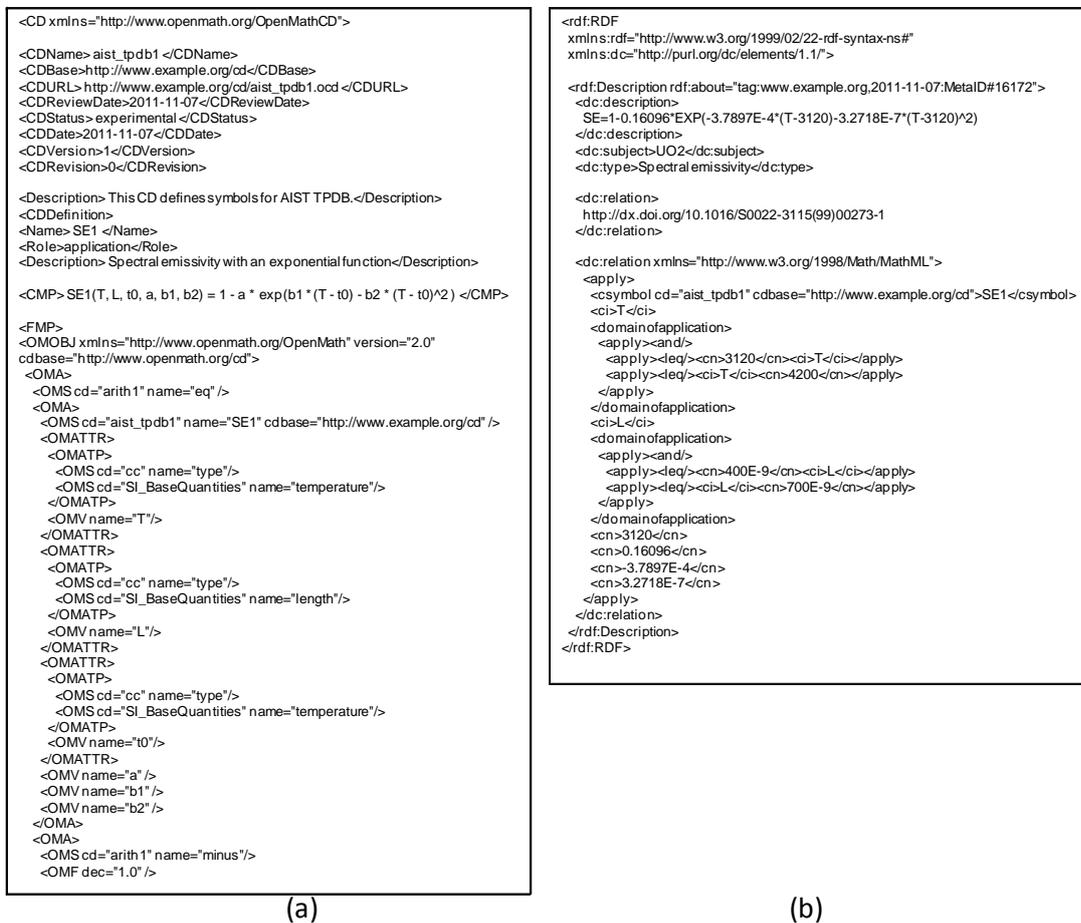


Figure 2. Part of the OpenMath entry for Eq. (1). (a) is a part of the CD definition for the generic equation format. (b) shows the RDF definition of a set parameter values (Fink, 2000).

The CD “aist_tpdb1” defines a function named SE1, which takes six arguments (Figure 2(a)): Temperature T, t0 (t0 is the lower limit temperature in which this function is valid, in this case, it is 3120K. T-t0 corresponds to ΔT in Eq. (1)) T0, (T-T0 corresponds to ΔT), and L (corresponding to wave length λ), are physical values. a1, b1 and b2 are coefficients of each term respectively. They are calculated from experimental data by a least square method. L is not

used in the expression, but it is defined in the CD as a required defining domain of the application of this expression. The “SI_BaseQuantities” content dictionary, which is provided by the OpenMath Society, is referred to in order to import the definitions of units and dimensions.

The RDF definition in Figure 2(b) gives values of parameters that are defined in the CD. They are estimated by the least square method and experimental data. The csymbol SE1 indicates the expression defined by http://www.example.org/cd_aist_tpd1. Also, it gives the type of material, UO2, doi of original article, and domain of application of the estimated values. Its definition is given in the MathML content markup, which is compatible with OpenMath markup.

The CD, which defines about 20 equations, and the RDF file, which defines about 60 sets of parameter values for estimating, fitting, and describing the thermophysical properties in the AIST database, are developed and verified. In these days, Web browsers, mathematical software packages, and open source tools support MathML, and there is no requirement for developing specific software tools to display and handle these data entries. Also, it is easy to extend, i.e., add new expressions, physical properties, and information, such as hyperlinks to other related information resources.

4 CONCLUSION

Mathematical knowledge is an important element of materials science and engineering. OpenMath provides a portable way to describe mathematical expressions and the content dictionary for the AIST thermophysical property database. It enables separating the generic definitions of equations from the individual set of parameter values. They can be shared on the Internet by referencing their URIs among many materials databases. Because many mathematical applications support OpenMath and MathML as their import and export data formats, this approach enables the exchange of mathematical knowledge among materials databases and other systems, such as mathematical systems and Web based systems. MKM (Mathematical Knowledge Management) (Farmer, 2004) is one research area within knowledge management that provides CDs of mathematical functions, operators, and units. We can utilize these resources to manage the empirical and theoretical equations of materials science and engineering.

5 ACKNOWLEDGEMENT

This research was supported by Grant-in-Aid for Scientific Research 23500131 from the Japan Society for the Promotion of Science (JSPS). The authors appreciate Mr. Toshiyuki Mizutani for his enormous technical support in designing and developing the OpenMath definition.

6 REFERENCES

- Ashino, T. (2010) Materials Ontology: An Infrastructure for Exchanging Materials Information and Knowledge, *Data Science Journal* 9, 54-61
- Baba, T., Yamashita, Y. & Nagashima, A. (2009) Function Sharing and Systematic Collaboration between a Networking Database System and Printed Media on Themophysical Properties Data, *J. Chem. Eng. Data*, 54 (9), 2745-2757
- Buswell, S, Capriotti, O., Carlisle, D.P., Dewar, M.C., Gaetano, M. & Kohlhase, M. (Eds.) (2004) The OpenMath Standard. Retrieved Sep. 19, 2012, from the World Wide Web: <http://www.openmath.org/standard/om20-2004-06-30/omstd20.pdf>
- Carlisle, D., Ion, P. & Miner, R. (Eds.) (2010) Mathematical Markup Language (MathML) Version 3.0. Retrieved Sep. 19, 2012 from the World Wide Web: <http://www.w3.org/TR/MathML3/>
- Farmer, W.M. (2004) MKM: a new interdisciplinary field of research, *ACM SIGSAM Bulletin*, 38(2), 47-52

Fink, J.K. (2000) Thermophysical properties of uranium dioxide, *Journal of Nuclear Materials*, 279, 1-18

Fukuyama, H. & Waseda, Y. (2009) *High-Temperature Measurements of Material, Advances in Material Research 11*, Berlin, German, Springer Berlin Heidelberg

Jasper, J.J. (1972) The Surface Tension of Pure Liquid Compounds, *The Journal of Information Science and Technology Association Journal of Physics and Chemical Reference Data*, 1(4), 841-1009

Ojala, T. & Over, H-H. (2008) Approaches in Using MatML As a Common Language for Materials Data Exchange, *Data Science Journal*, 7, 179-195

Yamaguchi, K. & Itagaki, K. (2002) "Measurement of High Temperature Heat Content by Drop Calorimetry", *Netsu Sokutei*, 29(4), 173-180 (in Japanese)

(Article history: Received 25 October 2012, Accepted 8 December 2012, Available online 22 December 2012)