

MATERIALS ONTOLOGY: AN INFRASTRUCTURE FOR EXCHANGING MATERIALS INFORMATION AND KNOWLEDGE

Toshihiro Ashino^{1*}

¹*Dept of Regional Development Studies, TOYO University, 2-36-5 Hakusan, Bunkyo-Ku, Tokyo, 112-0001, Japan*

E-mail: ashino@acm.org

ABSTRACT

We have rich information resources for materials science and engineering - raw measurement data, computational simulation methods, digitized handbooks, and digital libraries. However, these resources have a wide variety of formats, terminologies, and concepts, which makes it difficult to find appropriate information for materials design, development, and evaluation. One solution to this problem is to integrate these resources into a computer readable concept map, called a domain ontology, which describes concepts and relationships among the concepts in materials science and engineering. This paper describes a trial that constructs a standard of metadata description using ontology language and demonstrates the validity of this construction through data exchange among heterogeneous material databases. "Materials Ontology," which consists of several sub ontologies corresponding to substance, process, environment, and property, is developed using the ontology language of the Semantic Web, OWL, which enables the definition of a flexible and detailed structure of materials information. A versatile "materials data format" is built on the Materials Ontology as a component of the materials information platform and is applied to exchange data among three different thermal property databases, maintained by two major materials science research institutes in Japan.

Keywords: Ontology, Information Integration, Material Database, Materials Properties

1 INTRODUCTION

The scientific data and knowledge generated in organizations and institutes are available on the Internet as Wikis, e-Journals, and online databases. These resources are loosely linked with each other, but their representation, terminology, and formats are not standardized. In order to integrate such information resources distributed on the Web, constructing a domain ontology that represents a map of concepts and their relationships for a certain scientific realm is becoming a major concern of e-science. This ontology is needed to assist scientists/researchers in finding adequate information from these data resources and integrating scientific information for research and engineering (Rubin, Lewis, Mungall, Misra, et al., 2006; Nambiar, Ludaescher, Lin, & Baru, 2006).

For materials science in particular, there exist several trials to create a domain ontology: PLINIUS is a knowledge base that handles knowledge about ceramics research (van der Vet, Speel, & Mars, 1995); Ashino and Fujita (2006) offer a trial describing creep properties and a creep data analysis process with an ontology; and MatONT (Cheung, Drennan, & Hunter, 2008) is designed to support information integration for new materials research.

Ashino and Oka (2007) constructed a prototype of an information platform for materials information. It consists of two functions: one is an RSS metadata aggregation for materials databases and the other is a domain ontology for exchanging materials data among three heterogeneous materials databases that have been developed independently by AIST (National Institute of Advanced Industrial Science and Technology, Japan), NIMS (National Institute of Materials Science, Japan), and the materials data schema defined by MatDB, a fatigue and creep data schema designed by researchers in Japan. These databases consist mainly of the mechanical and physical properties of engineering materials.

Furthermore, as evaluation of materials from multiple viewpoints grows more important, we must evaluate materials not only for their cost performance but also for energy efficiency, environmental effects, and possible long term toxicity. This makes it even more important to integrate heterogeneous material databases. Materials Ontology enables the exchange of materials data among heterogeneous materials databases and other information resources.

2 DESIGN OF MATERIALS ONTOLOGY

Current Materials Ontology consists of the following seven ontologies, which are organized into three groups.

- ◆ Core Ontologies
 - Substance
 - Process
 - Property
 - Environment
- ◆ Materials Information
- ◆ Peripheral Ontologies
 - Unit Dimension
 - Physical Constant

Concepts related to materials science have been categorized into four groups of Core Ontologies: Substance, Process, Property, and Environment. These Core Ontologies give structured definitions of terms, names, and vocabularies for the basic concepts of each domain. Figure 1 shows part of the structure of the Property ontology, Property class, which corresponds to the fundamental data structure of materials property and has six properties (attributes). Three of them, PropertyName, UnitDimension, and UnitSystem, are ObjectProperties, which means they can take an instance of appropriate classes.

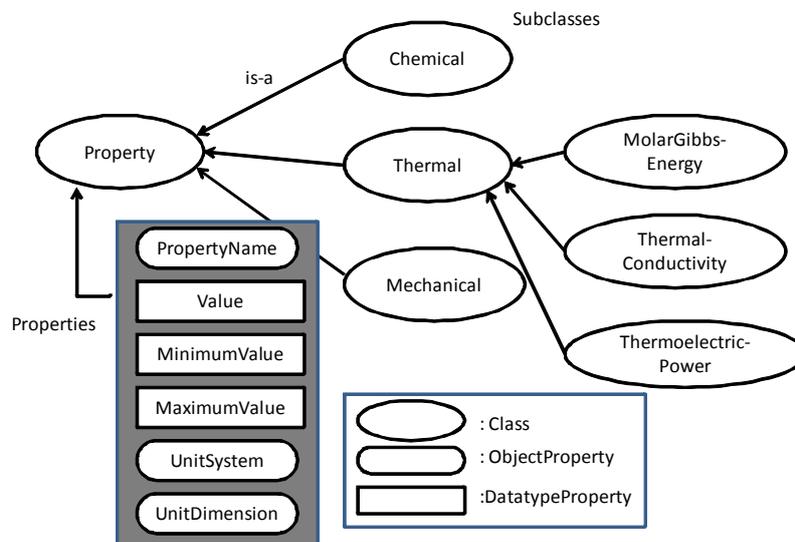


Figure 1. Structure of the Property Ontology. The Property class has the properties: PropertyName, Value, and UnitSystem. The subclasses, Chemical, Thermal, and Mechanical, correspond to the taxonomy of materials property.

Process ontology describes not only material processing, such as heat treatment, but also measurement methods and manufacturing processes, such as welding. Environment ontology describes the environmental parameters of materials processes, such as atmosphere, stress, and temperature. Substance ontology is a taxonomy of materials based on material types as shown in Figure 2. The designers of MatOnt have decided to construct a structure ontology because microstructure is very important information for advanced materials design. In this paper, however, the target databases are collections of data for engineering materials. Engineering materials, in many cases, are mixtures of several microstructures, and separated structure ontologies have not been developed. Environment ontology defines concepts related to the environmental parameters of testing, measurement, and processing materials, such as pH, Atmosphere, and HoldingTime.

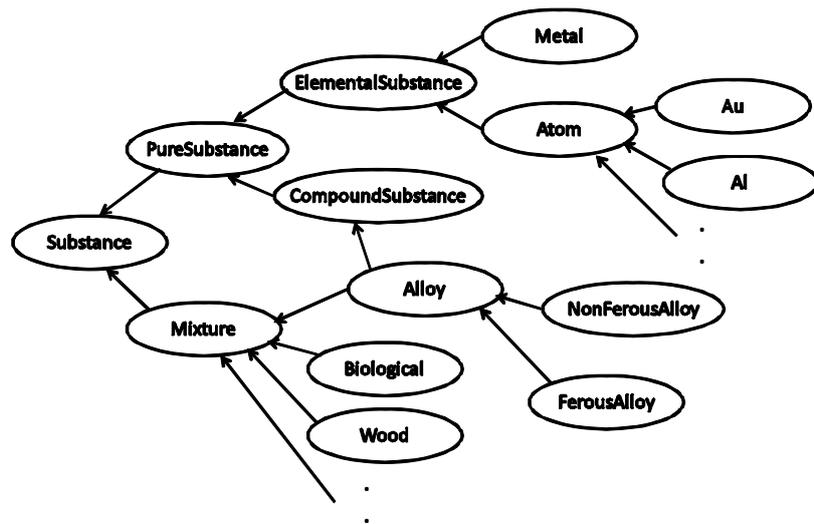


Figure 2. A part of the Substance ontology class network. The concept “Alloy” is a subclass of two classes, “CompoundSubstance” and “Mixture.”

Materials Information (Figure 3) is an aggregation of other classes corresponding to a record of materials data. Class EngineeringMaterial is an aggregation of ChemicalFormula, CommonName, SubstanceClass, and other class instances and data that are required to specify a single specimen. Class MaterialProperty is an aggregation of Environment, MeasurementMethod, Property, Equation, and Specimen. The ObjectProperty Specimen takes an instance of EngineeringMaterial as its value.

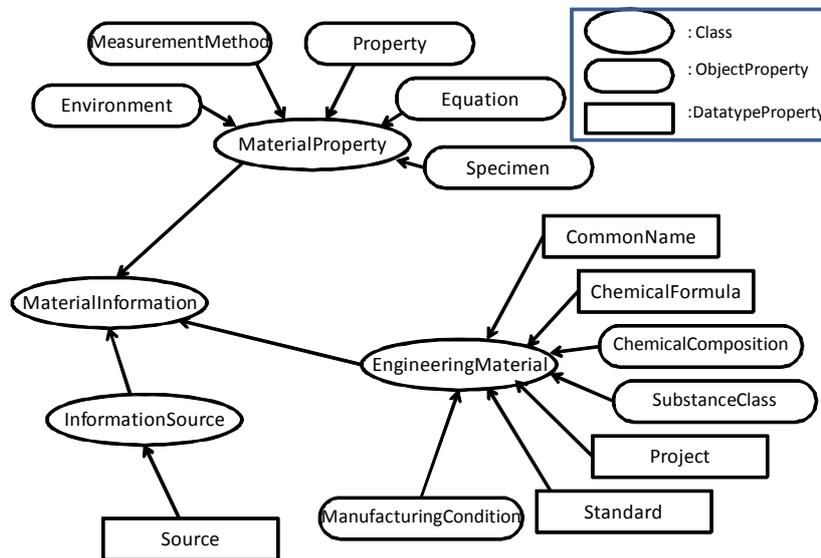


Figure 3. Data structure of MaterialInformation ontology. The MaterialInformation class aggregates material, its property, and data source. It corresponds to a single data entry in a materials database.

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:matinfo="http://codata.jp/OML-MaterialInformation.owl#"
  xmlns:substance="http://codata.jp/OML-Substance.owl#"
  xmlns:property=http://codata.jp/OML-Property.owl#
  xml:base="http://codata.jp/AI.owl">
  <owl:Ontology>
    <owl:imports rdf:resource="http://codata.jp/OML-MaterialInformation.owl"/>
  </owl:Ontology>
  <matinfo:MaterialProperty rdf:ID="MaterialProperty1">
    <matinfo:Specimen>
      <matinfo:EngineeringMaterial rdf:ID="EngineeringMaterial">
        <matinfo:CommonName>Aluminum</matinfo:CommonName>
        <matinfo:ChemicalComposition>
          <substance:Al rdf:ID="Al"/>
        </matinfo:ChemicalComposition>
      </matinfo:EngineeringMaterial>
    </matinfo:Specimen>
    <matinfo:Property>
      <property:ThermalConductivity rdf:ID="ThermalConductivity1">
        <property:UnitSystem rdf:resource="http://codata.jp/OMLUnitDimension.owl#SI"/>
        <property:Value rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
          297
        </property:Value>
      </property:ThermalConductivity>
    </matinfo:Property>
  </matinfo:MaterialProperty>
</rdf:RDF>

```

Figure 4. Thermal conductivity data of pure aluminum specimen written in Materials Ontology

Figure 4 shows an example data entry written in Materials Ontology describing the thermal conductivity of pure aluminum. The first third defines the namespaces and imports the required ontology files. Data entry is an instance of MaterialProperty class, consisting of instances of Specimen class and Property class. They specify the name and chemical composition of the specimen and the name, value, and measurement unit system of the physical property.

Peripheral Ontologies describe more generic concepts for scientific information, UnitDimension and Physical Constant. The UnitDimension ontology design is based on MathML syntax (Harder & Devitt, 2003), in order to write the unit conversion or other operations in MathML. This enables Materials Ontology to link with digital libraries of empirical equations, theoretical equations, and data analysis equations for materials science written in MathML. The URIs for basic dimensions and units are defined in codata.jp namespace (Table 1).

Table 1. Example of URIs allocated for units and dimensions

Length	http://codata.jp/OML-UnitDimension.owl#length
MagneticFlux	http://codata.jp/OML-UnitDimension.owl#magneticFlux
Mass	http://codata.jp/OML-UnitDimension.owl#mass
PlaneAngle	http://codata.jp/OML-UnitDimension.owl#planeAngle
Power	http://codata.jp/OML-UnitDimension.owl#power
Acceleration	http://codata.jp/OML-UnitDimension.owl#acceleration
CurrentDensity	http://codata.jp/OML-UnitDimension.owl#currentDensity

3 IMPLEMENTATION AND AN APPLICATION OF MATERIALS ONTOLOGY FOR MATERIAL DATA EXCHANGE

In order to utilize software tools and to maintain compatibility with other ontologies, W3C (the World Wide Web Consortium) recommends the use of OWL, the Web Ontology Language (Bechhofer, Harmelen, Hendler, Horrocks, McGuinness, Patel-Schneider, et al., 2004), to implement Materials Ontology. OWL is a component of the so called Semantic Web framework, a collection of standardized definition methods of data structures, ontologies, rules, and logic that share data and knowledge on the Internet. One of the applications of the Semantic Web shares scientific knowledge and information (Marshall & Prud'hommeaux, 2008).

We used the Protege ontology editor (Knublauch, Ferguson, Noy, & Musen, 2004) to edit and visualize and defined about 600 classes. From the huge amount of conceptual material related to materials properties, we have focused on thermal properties because they are found in all three - AIST, NIMS, and MatDB - materials databases.

Data exchange among three heterogeneous databases is implemented with an intermediate data representation based on Materials Ontology. Figure 5 displays part of thermal property data structures defined in the NIMS, Materials Ontology, and AIST databases. The gray arrows indicate corresponding concepts. For example, the value of the thermal conductivity is defined as a tensor (k11, k22, k33), in the AIST database; the "field" accepts both tensor and scalar; and the corresponding value class in Materials Ontology accepts a scalar value and arbitrary size matrix.

These two schemas show typical styles of material data representation with relational data models. The NIMS data structure is a straightforward derivation of an experimental data structure; each data field has the name of its contents, e.g., thermalConductivity and chemicalFormula. In contrast, the AIST data schema is designed for extensibility. Its fields have generalized names, e.g., "property." In this field, the names of materials properties are stored as string data. This design enables the addition or modification of property names after the schema has been fixed.

The corresponding data fields are mapped into the Materials Ontology network, using an XSLT (Clark, 1999) template to translate. Although the template synthesis can be automated, it is still done manually.

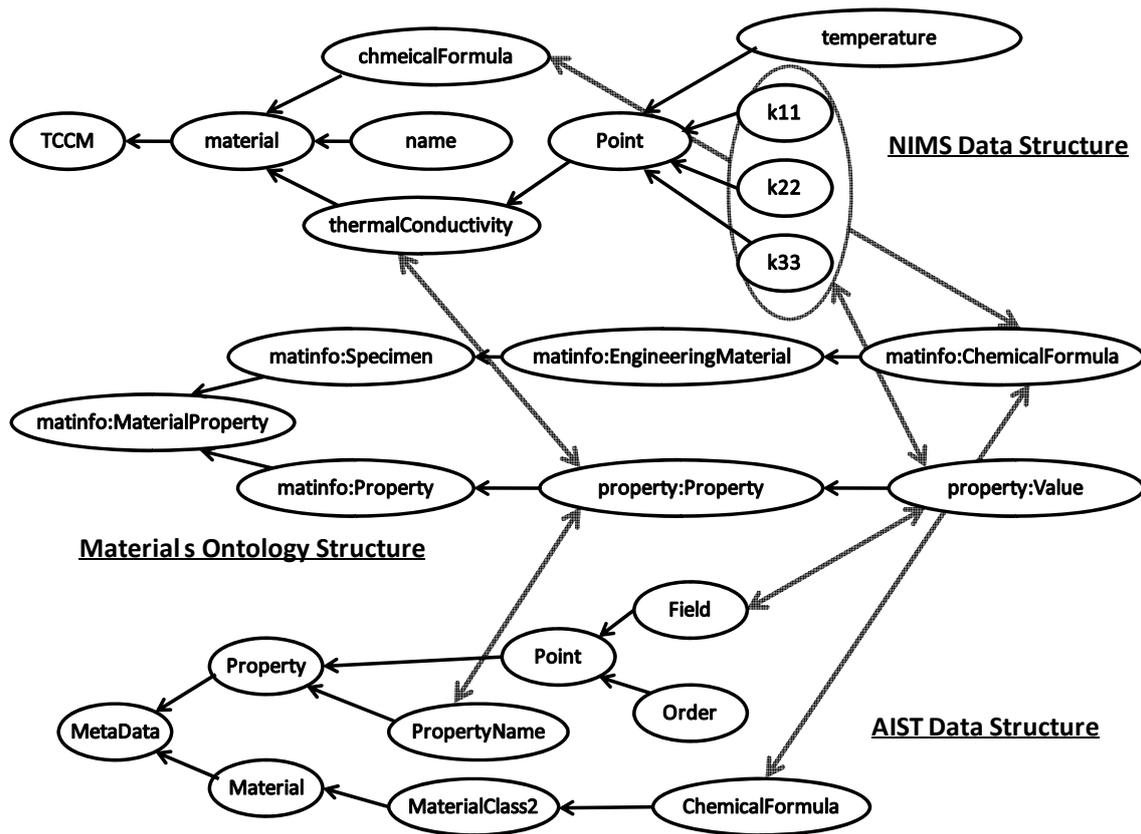


Figure 5. Data schema structures for thermal conductivity of the NIMS, Materials Ontology, and AIST databases and correspondence of their data fields; “matinfo:” is the prefix of the material information ontology.

4 DISCUSSION

There are several trials that have developed a materials database format with XML, such as MatDB and MatML (Kaufman & Begley, 2003). However, MatML has no definition of concepts. In other words, it has no semantics. It takes an approach similar to that of the AIST materials database discussed in the previous section, which achieves generality. Of course, it is very difficult to agree on definitions of concepts. In order to exchange data among heterogeneous materials databases, however, some kind of translation table for materials data is essential. It is reasonable to start extracting common concepts from existent databases or handbooks and gradually extend this work. Modularity and extensibility should be considered in the basic structure. In OWL, they are provided mainly by namespaces and prefix URI, which should be selected carefully.

Ashino and Fujita (2006) provide an example of materials science descriptive knowledge using OWL. The Materials Ontology gives essential common terms or notations for this kind of knowledge base. Another possible component is a digital library of empirical and theoretical equations for materials science knowledge written in MathML, the mathematical markup language. These three possibilities use a common data format, XML, and can work together on the Web.

5 CONCLUSION

Materials Ontology, which covers concepts of materials substance, property, environment, and process, is developed and applied for data exchange among heterogeneous materials databases. It serves as a base with which to integrate databases and other materials information resources, such as computational simulations, equation solvers, and documents.

Materials Ontology is a structured common dictionary of concepts related to materials science. It is used to describe materials science knowledge within a Semantic Web framework, mapping concepts among heterogeneous materials data resources, and make an intelligent search of materials information, data, and knowledge.

Documentation of Materials Ontology is available from our materials information portal, <http://musigny.rds.toyo.ac.jp:8080/>, currently available only in Japanese but in the process of translation into English.

6 ACKNOWLEDGEMENTS

Special thanks to Dr. Nobuto Oka (Aoyama Gakuin University), who authored the OWL definitions, and Dr. Mitsutane Fujita (former of NIMS), who provided knowledge about the structure of materials information.

This research work is supported by the NEDO (New Energy and Industrial Technology Development Organization in Japan) project “Research and Development to Promote the Creation and Utilization of an Intellectual Infrastructure” fy2006-2007.

7 REFERENCES

- Ashino, T. & Fujita, M. (2006) Definition of Web Ontology for Design-Oriented Material Selection. *Data Science Journal* 5, 52-63.
- Ashino, T. & Oka, N. (2007) Development of Information Platform for Data Exchange between Heterogeneous Material Data Resources. *Proceedings of PV 2007 Conference, Ensuring the Long-term Preservation and Value Adding to Scientific and Technical Data*, Munich.
- Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein L. A. (2004) OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004. Retrieved April 1, 2009 from the World Wide Web: <http://www.w3.org/TR/owl-ref/>
- Cheung, K., Drennan, J., & Hunter, J. (2008) Towards an Ontology for Data-driven Discovery of New Materials. *AAAI Workshop on Semantic Scientific Knowledge Integration*, Stanford University, 26-28.
- Clark, J. (Ed.) (1999) XSL Transformations (XSLT) Version 1.0, W3C Recommendation 16 November 1999, Retrieved April 1, 2009 from the World Wide Web: <http://www.w3.org/TR/xslt>
- Harder, D. W. & Devitt, S. (Eds.) (2003) Units in MathML, W3C Working Group Note 10 November 2003. Retrieved April 1, 2009 from the World Wide Web: <http://www.w3.org/TR/mathml-units/>
- Kaufman J. G. & Begley E. F. (2003) MatML: A data interchange markup language. *Advanced Materials and Processes* 161(11), 35-36.
- Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004) The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. *Third International Semantic Web Conference*, Hiroshima, Japan, 229-243.
- Marshall, M. S. & Prud'hommeaux, E. (Eds.) (2008) A Prototype Knowledge Base for the Life Sciences, W3C Interest Group Note 4 June 2008. Retrieved April 7, 2009 from the World Wide Web: <http://www.w3.org/TR/2008/hcls-kb/>
- Nambiar, U., Ludaescher, B., Lin, K., & Baru, C. (2006) The GEON portal: accelerating knowledge discovery in the geosciences. *WIDM '06: Proceedings of the 8th annual ACM international workshop on Web information and data management*, 83-90.

Rubin, D. L., Lewis, S., Mungall, C., Misra, S., Monte, Westerfield, M., Ashburner, M., Sim, I., Chute, C., Solbrig, H., Story, M-A., Smith, B., Day-Richter, J., Noy, N. & Musen, M. (2006) National Center for Biomedical Ontology: Advancing Biomedicine through Structured Organization of Scientific Knowledge. *A Journal of Integrative Biology* 10(2), 185-197.

van der Vet, P. E., Speel, P-H., & Mars, N. J. I. (1995) Ontologies for very large knowledge bases in materials science: A case study. *The Second International Conference on Building and Sharing Very Large-Scale Knowledge Bases*, University of Twente, 73-83.

(Article history: Received 27 October 2009, Accepted 18 May 2010, Available online 26 June 2010)