Faceted Framework for Metadata Interoperability

메타데이터 상호운용성 확보를 위한 패싯 프레임워크 구축

Seungmin Lee*

ABSTRACT

In the current information environment, metadata interoperability has become the predominant way of organizing and managing resources. However, current approaches to metadata interoperability focus on the superficial mapping between labels of metadata elements without considering semantics of each element. This research applied facet analysis to address these difficulties in achieving metadata interoperability. By categorizing metadata elements according to these semantic and functional similarities, this research identified different types of facets: basic, conceptual, and relational. Through these different facets, a faceted framework was constructed to mediate semantic, syntactical, and structural differences across heterogeneous metadata standards.

초 록

현재의 정보환경에서 메타데이터 상호운용은 정보를 관리하고 조직하는 데 있어 유용한 도구가 되고 있다. 하지만 상호운용성 확보를 위한 현재의 방법들은 메타데이터 요소의 의미를 고려하지 않은 피상적인 매핑에만 중점을 두고 있다. 본 연구에서는 이러한 문제들을 해결하기 위해 패싯분석을 적용하여 메타데이터 상호운용성을 확보하고자 한다. 메타데이터 요소들의 의미와 기능적인 유사성에 따라 요소들을 분석하여, 기본패싯, 개념패싯, 관계패싯의 세 가지 유형의 패싯을 확인하고, 이를 통해 의미적, 구문적, 구조적 상호운용성을 제공할 수 있는 패싯 프레임워크를 구축하였다.

Keywords: metadata, metadata interoperability, facet analysis, basic facet, conceptual facet, relational facet, faceted framework
메타데이터, 메타데이터 상호운용, 패싯분석, 기본패싯, 개념패싯, 관계패싯, 패싯 프레임워크

* School of Library and Information Science, Indiana University(seungmin@umail.iu.edu)

▪ Received : 15 May 2010 ▪ Revised : 4 June 2010 ▪ Accepted : 15 June 2010

▪ Journal of the Korean Society for Information Management, 27(2): 75-94, 2010.
〔DOI:10.3743/KOSIM.2010.27.2.075〕

www.kci.go.kr

1. Introduction

In the current information environment, a metadata standard is a powerful tool for describing and managing resources and for standardizing resource description. However, as the number of heterogeneous metadata standards has increased to satisfy the specific needs of different communities, it has become more and more complicated to manage information resources. The flood of purpose-specific standards is generally incompatible with other standards, and this heterogeneity of metadata standards has limited interoperability among standards: Each of these domain-specific standards exists independently of other standards and is only meaningful (and useful) for the purposes of a specific community.

Efforts have been undertaken in the library community to overcome the problems of multiple metadata standards. These efforts have generated several different approaches to minimizing differences between heterogeneous standards and achieving or improving metadata interoperability. However, in spite of their strengths, each approach has weaknesses that limit complete metadata interoperability.

The ideal approach would be a uniform metadata framework: If all communities were required to use the same standard, problems of metadata interoperability would be resolved. However, this approach would only be viable if implemented before standards were adopted by different communities (Chan 2005). Among the other approaches that have been proposed, element mapping and crosswalks are simple approaches based on the establishment of element relationships. Although they are relatively easy to develop and do not require any specific format or model, they are not capable of achieving metadata interoperability because they focus on element-level or incomplete schema-level mappings, which can lead to inappropriate mapping results. The application profile specifies a set of metadata terms to be used in a metadata application and imposes specific element relationships however, the dependence of this approach on element-level mapping can undermine reliable interoperability. The use of application profiles can also be an obstacle to cross-community interoperability and could ultimately lead to the development of a plethora of application profiles. The metadata registry is currently the most powerful approach to metadata interoperability because it encompasses the strengths of other approaches without many of the drawbacks. However, a registry requires a long-term commitment to development and an organization that will take full responsibility for the registration and authentication of metadata elements. As with the application profile, there is also the potential for development of heterogeneous metadata registries with varying structures, which would further complicate problems of metadata interoperability.

All of these approaches are intended to eliminate or minimize heterogeneity between different metadata standards in order to make them interoperable with each other. However, the limitations of these approaches are serious obstacles to achieving reliable metadata interoperability and may actually diminish the flexibility and utility of any interoperability that is achieved. Finally, there is a serious weakness

www.kci.go.kr

shared by all of these approaches: Although each approach provides some level of support for interoperability, represented either as a simple table or a complicated model, the results have generally not been evaluated or tested in a formal way. Thus, no one approach can claim that it is, in fact, a reliable approach to metadata interoperability.

The goal of this research is to discover a practical approach that will make such an approach more viable for metadata interoperability. This paper investigates the method for constructing data model for metadata interoperability with the application of facet analysis.

2. Research Description

One of the main factors limiting interoperability is that many standards have been designed to satisfy the needs of a particular community or customized for a specific purpose. The nature of these standards often prevents their use in other communities. Procedural problems also obstruct the achievement of interoperability. The typical process for establishing metadata interoperability is to extract selected elements and identify relationships between them. However, the selection of elements limits the breadth of metadata interoperability and may even make problems of interoperability worse.

The structure generated by individual approaches may also be problematic in achieving metadata interoperability. In many cases, an element in one standard may be related to two or more elements in another standard based on semantic similarity. Interoperability should support these multiple relationships and indicate the level of granularity associated with each. However, current approaches tend to generate a flat relational structure that may not be flexible or systematic enough to represent complex relationships. Such structural weaknesses have caused problems with the reliability of interoperability (e.g., missing elements and/or ambiguous element relationships).

Another limitation is a concentration on representational meanings that does not take into account the structural differences between standards. Metadata standards may have different structures that affect the meaning of elements: An element frequently has a specific meaning inherited from its domain context and from the relational structure of the standard, but this meaning is often obscured by the element's superficial meaning. Due to different contexts and different levels of granularity, the semantics of two elements may be very different even when they share the same label. Although interoperability approaches should consider such differences, all too frequently they focus on the similarity of element names to establish element relationships, resulting in the identification of ambiguous, misleading, or inaccurate relationships. When these limitations are not addressed, they present significant obstacles to achieving metadata interoperability.

These considerations address the weaknesses and limitations that confront current approaches to interoperability, which is key to achieving successful interoperability across heterogeneous standards. Because the problems of interoperability are closely tied to

www.kci.go.kr

the purpose-specific or domain-specific characteristics that lead to inconsistencies in interpretation across standards, the key to overcoming these difficulties is the elimination of inconsistencies across different standards. This research investigates the criteria and processes for achieving metadata interoperability in order to develop a new approach that meets the following criteria: It is sophisticated and comprehensive enough to encompass existing metadata standards and their applications; it represents relationships across standards clearly; and it consistently makes standards interoperable.

3. Research Approach

3.1 Overview

To make metadata standards interoperable, it is necessary to understand the semantics of each standard as reflected in the components of the standard: the elements, the syntax, and the structure. Without understanding the inherent semantics of metadata standards, interoperability is merely a superficial mapping based on the labels of elements. In contrast, this research considers metadata interoperability from a semantic point of view that requires more than simple agreement on the static meaning of elements. It entails the coordination of meanings to mediate not only content-related but also syntactical and structural differences between standards.

The research proposed here focuses on the development of a faceted system that provides for the organization of elements by splitting (separating) semantically unrelated or dissimilar meanings and lumping (grouping) related or similar meanings (Yang 2004). By concentrating on the representation of these semantic relationships, the research sought to develop a faceted framework that could represent the semantics of diverse standards, systematically organize the components of standards, and facilitate integration of the semantics of individual metadata standards.

3.2 Theoretical Foundations

Interoperability is a broad concept whose meaning varies according to the context or purpose to which it is applied. Generally, interoperability is defined as “the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system” (Moen 2003). However, the meaning of interoperability must be further specified in relation to metadata standards.

Many definitions of metadata interoperability emphasize the effective sharing and reuse of information through removal of the heterogeneous characteristics of individual metadata standards. More specifically, metadata interoperability refers to the ability to exchange and interpret data across two or more metadata standards by minimizing syntactic, structural, and semantic inconsistencies and thereby maximizing opportunities for sharing and reusing information (Blanchi and Petrone 2001; Miller 2000). Central to this definition is the potential for metadata stand-

www.kci.go.kr

ards to cross the boundaries between different metadata contexts by removing any heterogeneity, thus facilitating the unambiguous exchange of metadata records (Arms et al. 2002; Oltmans 2001, 2).

Making metadata standards interoperable is ultimately an effort to integrate heterogeneous standards. The key is to establish consistency across standards (Gill and Miller 2002). However, each metadata standard consists of many components (i.e., elements, syntax, and structure) which do not function independently but are interconnected and operate together. To achieve fully functional metadata interoperability, consistency must be achieved at the component level (i.e., between a component of one standard and the corresponding component(s) of other standards). Accordingly, this research narrowed the meaning of metadata interoperability to component-level interoperability. By focusing on the components of descriptive metadata standards that have corresponding components in other standards, the research focused on establishing relationships between components to achieve semantic interoperability across metadata standards.

Facet, in its simple meaning, is a conceptual categorization. It generally refers to a concept group, consisting of generic terms, used as general manifestation of a compound subject to denote component of the subject (Ranganathan 1967). Each facet contains a number of terms that will be considered to be conceptually equivalent. In the library community, facet is defined as “clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject” (Taylor 1992). From slightly different perspectives, facet also refers to a partitioning of vocabulary or grouping of terms obtained by the division of a subject discipline into homogeneous or semantically cohesive categories (Svenonius 2000).

The process of partitioning domain vocabulary and generating facet is often called facet analysis (or faceting). It is a mental process involving the analysis of a subject into its facets based on a set of postulates and principles, resulting in a knowledge structure with clearly delineated semantic relationships between concepts. This structure provides a framework to accommodate various types of concepts along with syntax rules for their combination (Kumar 1987).

Facet analysis derives two processes: analysis and synthesis. Ranganathan (1967) demonstrated analysis as “the process of breaking down subjects into their elemental concepts.” These concepts can be synthesized, which is the process of recombining concepts into subject strings or creating new compound terms (Ranganathan 1967).

Through the process of analysis and synthesis, facet analysis can be used as a tool to identify and represent relationships between concepts in a certain subject. Based on the concept relationships, it also provides a framework of vocabulary with enough flexibility to include new subject in it because it can synthesize or combine subject with facets. Facet analysis also provides for the organization of concepts in modular hierarchies by separating unrelated or dissimilar concepts and grouping related or similar concepts. Thus, relevant concepts can be identified

www.kci.go.kr

by partitioning domain vocabulary into mutually exclusive facets (Priss and Jacob 1998).

In spite of these strengths, facet analysis has weaknesses in organizing and categorizing concepts. The biggest problem is that it is difficult to define facets and to prescribe the semantic range of each facet. Because of the ambiguous or abstract semantics of facets, there are concepts that cannot be categorized exclusively into only one facet (Svenonius 2000). Another problem is raised when two concepts with different attributes have the same label. Facet analysis does not provide a way to distinguish between these concepts, resulting in the ambiguity of concept relationships.

To complement these weaknesses and make the meaning of each facet clear, this research divides facets into three types according to their functions: basic, conceptual, and relational facet. Although the conceptual and relational facets are related to the basic facet, each has different functions according to their functional roles when describing or indicating specific aspects of the basic facet. By applying these different types of facets, semantic similarity between concepts can be clearly identified and the subtle meaning of each facet can be reflected in the facet structure. In addition, the levels of similarity between concepts can be represented according to the assigned facets under the same basic facet. Another strength is that these different types of facets can support consistent extension of facet structure by representing concept relationships.

4. Faceted Approach to Metadata Interoperability

4.1 Selecting Metadata Standards

Because the faceted framework was to be constructed based on the analysis of existing metadata standards, the first step was the selection of metadata standards. Three criteria were used to limit the range of possible standards: (1) the standard must be currently in use in the library community; (2) the standard must not be limited to a specific type of resource but must be able to deal with many types of resources; and, (3) the standard must have been designed for general use rather than for a specific context. Based on these criteria, two standards were selected: MAchine Readable Cataloging (MARC) and Metadata Object Description Schema (MODS).

MARC has been the accepted bibliographic standard in the library community for over 40 years. It provides a standardized and machine-readable structure for encoding and exchanging bibliographic data. Using this structure, libraries can organize and store bibliographic data in a consistent and standardized way that provides a basis for cooperative cataloging systems. Given these strengths, other institutions that deal with information resources (e.g., archives and museum) have also adopted the MARC format for organizing and managing resources. Although MARC originated as a means to communicate bibliographic data about print materials, it has evolved to facilitate the description of a broad range of resources, including computer files, maps, serials, music, and audio-visual

www.kci.go.kr

materials (Moen and Benardino 2003).

MODS is a MARC-compatible XML standard that establishes a simplified set of conventions for encoding descriptive data. It is designed to be able to convert selected data from existing MARC21 records as well as to enable the creation of original records. It is particularly appropriate for representing digital resources that require rich descriptions, but it is not as complex as MARC, making MODS records easier and quicker to create than MARC records. MODS uses language-based tags for data elements that have been derived from MARC21, which makes MODS more user-friendly because, in contrast to the numeric tags of traditional MARC, tags can be easily understood by anyone dealing with the raw metadata record. With these strengths, MODS is becoming an accepted metadata standard that can support simple metadata representations as well as more complex metadata records and is flexible enough to deal with a variety of resource types.

Although MODS has a high level of compatibility with MARC systems and most of the elements defined in MODS having equivalents in the MARC21 format (Guenther 2003, 139), the previous mappings between these two systems could not allows for reliable conversion because of the disparity of granularity and semantic differences between them. As a result, multiple MARC elements are indicated for a single MODS element and some specificity would be lost during the mapping process. Given this situation, this research intended to construct a framework that identifies semantic relationships and achieves semantic interoperability between MARC and MODS by applying facet analysis.

4.2 Generating Vocabulary of Metadata Elements

Once metadata standards have been selected, the next step is to generate vocabulary of metadata elements by extracting elements from different standards. Construction of the element vocabulary began with an analysis of the MARC and MODS metadata standards in order to extract elements and their labels.

Each element has unique meaning to represent certain aspect of a resource, such as author, title, and subject. However, the meaning is not just from the aspects of the resource, but also from the context that is usually reflected in the structure of each metadata standard. This context is important when considering metadata interoperability because it affects on the semantics of elements and often makes the same element to be used in different ways, even when they have the same definitional meaning. To mediate this potential inconsistency, the context of each standard, including its structure and syntax, was considered during the analysis. Thus, when extracting elements, it is necessary to identify the semantics of each element in its original context reflected in the structure and syntax of the standard.

A metadata standard usually has two general levels at which elements occur in the structure. Main elements occur at the top of the structure and usually have broad meanings that establish the descriptive range of the standard (e.g., author, title, subject). Main elements can be used either as actual elements

www.kci.go.kr

that describe certain aspects of a resource or as superordinate semantic categories that establish relationships among the subordinate elements that constitute the second general level in the hierarchy. The meaning of a main element is actually specified by its more detailed, subordinate elements, which divide the broad meaning of the superordinate element in order to describe specific aspects of a resource.

Although the syntax of the standard also guides how the meaning of an element is to be interpreted, there are also critical degrees of deflection in the range and number of elements because of the disparity of granularity that occurs across standards. As a result, elements with the same meaning can occur as a superordinate element in one standard or a subordinate element in another standard. This contextual difference across standards was also considered in order to clearly identify the semantics of elements.

Guided by these considerations, the first step in creating the element vocabulary was the extraction of superordinate elements from the MARC and MODS standards in order to analyze their inherent meanings and identify the semantic range of each standard. MODS has 20 top-level elements and 55 sub-elements. The superordinate elements in MODS can have actual value(s) that indicate specific aspects of a resource or they can function as category labels encompassing subordinate elements to which actual values are assigned. In contrast, MARC has 10 groups of fields, each of which has no specific value but functions as a category to group the relevant MARC components (i.e., fields). Among MARC's 10 groups of fields, only eight are actually used to represent specific aspects of a resource. The small number of groups of fields results in broad and often ambiguous interpretations of the semantics of fields. In order to capture more detailed meanings, MARC fields (e.g., 100, 110, and 130) were treated as superordinate elements because they have functions similar to the superordinate elements in MODS. Given this situation, MARC was treated as a single-layered structure comprised of multiple fields. A total of 221 MARC fields were extracted for analysis as superordinate elements.

After extracting elements from the selected standards, elements were compared to identify those which were semantically common to those standards. However, the meanings of elements were expressed in different formats according to the syntax of the original standard, their differing structural positions, and varying degrees of deflection and disparity of granularity. To mediate lack of semantic regularity, the analysis focused on the basic meanings of elements regardless of variations in representation. The element's structural position in each standard was also considered to mitigate structural and syntactical differences among elements. The result was identification of 13 shared meanings that were commonly used in the two standards (See Table 1).

Each shared meaning functions as a core element that connects different standards through the subordinate elements nested under that meaning. The common meaning also serves as a semantic category in that it is used to collocate and store superordinate and subordinate elements extracted from the different standards. After identifying the set of 13 shared mean-

www.kci.go.kr

〈Table 1〉 List of 13 Common Elements Used across the MARC and MODS

Shared meaning	MARC	MODS
title	2XX	〈title〉
publisher	260	〈originInfo〉
creator	100, 700, 710, 711, 720, 800	〈name〉
language	041, 546	〈language〉
date	260, 033, 046	〈originInfo〉
relation	76X, 77X, 78X	〈relatedItem〉
identifier	010, 020, 022, 024, 028, 037, 074, 856	〈identifier〉
resource location	852, 856	〈location〉
description	300, 5XX	〈physicalDescription〉
copyright	017, 506, 540, 542	〈accessCondition〉
format	007, 256, 300, 533	〈physicalDescription〉
type	008, 047, 655	〈typeOfResource〉 〈genre〉
subject	600, 610, 611, 630, 648, 650, 651, 653, 654, 656, 662, 752	〈subject〉

ings, extracted subordinate elements were collocated based on semantic similarity and were grouped under the corresponding shared meaning (See Table 2). This vocabulary shows the semantic range of the selected standards and functions as the foundation of constructing the faceted framework.

4.3 Concept Group Identification

After the vocabulary of metadata elements had been created, semantic relationships between elements in the vocabulary were established. Establishing semantic relationships required analyzing the core meaning of each element and identifying shared referents across elements that could be potential facets. Because of the differing contexts and the disparities of granularity across standards, however, significant differences might still exist between elements, even when semantic categories with common meanings had been identified and elements from the different standards grouped

〈Table 2〉 Example of Element Vocabulary for the Category Title

Category	MODS	MARC
Title	〈titleInfo〉 〈title〉 〈subTitle〉 〈partNumber〉 〈partName〉 〈nonSort〉 〈abbreviated〉 〈translated〉 〈alternative〉 〈uniform〉	130, 210, 240, 242, 243, 245, 246, 247, 490, 730, 740, 793, 830

www.kci.go.kr

under appropriate categories.

To address this possibility, it was necessary to extract the basic meaning of each element based on the definitions of the elements from the standards where it appears. This process removes semantic differences between two or more similar but ambiguous elements by identifying their core meanings and translating the elements into concepts and is therefore key to achieving semantic interoperability.

Facet analysis was applied to analyze elements during extraction by breaking them into their constituent units of meaning and translating them into the concepts that made up the semantics of each element. This extraction and translation process identified concepts and the relationships between concepts. At the same time, facet analysis generated a vocabulary of concepts that clarified the specific meaning of each element and mediated semantic difference between elements.

In creating the vocabulary of concepts, elements in the 13 common meanings in the vocabulary of elements were analyzed to identify its basic units of meaning. The facet analysis was then used to extract the basic meanings and to translate each meaning as a concept. Through the processes of extraction and translation, the semantics of each element were clearly demonstrated and all elements representing the same concept were identified and collocated under the shared meaning, regardless of different labels.

Because a concept translated through facet analysis was not changed even when the element was presented in different formats, relationships among elements could be clearly established by connecting elements from different standards through the shared meaning. In addition, because meaning, structure, and syntax were all involved in the analysis of concepts, structural and syntactical implications of a standard were also reflected in the concept relationships. In understanding this research, it was assumed that, if elements from different standards referred to the same concept, they could be identified as interoperable regardless of their originally heterogeneous formats. Table 3 demonstrates how the elements in the category Title were converted into a set of concepts.

〈Table 3〉 Example of Facet Vocabulary for the Category Title

Category	MODS	MARC	Facet
Title	〈titleInfo〉 〈title〉 〈subTitle〉 〈partNumber〉 〈partNarne〉 〈nonSort〉 〈abbreviated〉 〈translated〉 〈alternative〉 〈uniform〉	130, 210, 210, 240, 242, 243, 245, 246, 730, 740, 793, 830	title main title alternate title subtitle part part number part name abbreviated title translated title alternative title uniform title

www.kci.go.kr

The translated concepts for each element in the vocabulary of concepts were considered as facets which represent the core meaning of each element in a context-independent form. Because the facets functions as a surrogate for the meaning of that element, the vocabularies of elements and concepts provide guidelines that allow different metadata elements to be categorized according to their semantic similarities. They also eliminate any representational ambiguity resulting from the different contexts of individual standards as reflected in their syntaxes and structures. Thus, facet functions as a semantic category of concepts extracted from metadata elements and provides conceptual space to encompass semantically related concepts.

4.4 Construction of Faceted Framework

Although the element-to-concept conversion process supports identification of semantic commonality and mediates differences across elements from different standards, differences may still exist between concepts. Even if the same facet is assigned to a set of elements and concepts, it may not be easy to link them because of differences in their semantic range, resulting in that some concepts cannot be categorized exclusively into only one facet (Svenonius 2000). Another problem is raised when two concepts with different attributes have the same label. Facet analysis does not provide a way to distinguish between these concepts, resulting in the ambiguity of concept relationships. One way to avoid ambiguity is to establish relationships between facets according to their functional similarities. This relationship can be represented as a conceptual structure that provides information about how a facet is tied to other concepts and thus creates a faceted framework. To address this ambiguity and to establish functional relationships between facets, this research divides facets into three types according to their functional roles: basic, conceptual, and relational facet.

Basic facet refers to fundamental attribute to every entity in a subject. The meaning of each basic facet will be broad and abstract in order to encompass related facets in the subject. By encompassing those related facets, it prescribes the semantic range of facet structure generated by facet analysis. Each basic facet is located at the top of the structure without any superordinate facet and functions as a starting point of the deployment of the structure. Therefore, basic facet provides a semantic framework in which related facets in a subject can be placed together.

Conceptual facet is representative category of the attributes of concepts (Jung, Sung, and Park 2006). It can be considered as subordinates of basic facets. Although basic facet does not include concepts extracted from elements and only encompass related subordinate facets, conceptual facet categorizes actual concepts. Thus, this type of facet functions as category of concepts. It is located under each basic facet and constructs hierarchical facet structure by specifying the semantic range of the basic facets. This facet also defines representative aspects, properties, and characteristics of a concept in a subject.

In contrast to conceptual facet that describes specific

www.kci.go.kr

aspects of basic facet, relational facet is used to semantically encompass different kinds of the basic facets that, taken together, constitute the basic concept.

Construction of a faceted framework was based on the semantic categories of concepts associated with a subject and reflected in differences of functional granularity between facets. Concepts derived from superordinate elements can have a broad range of meanings: Even though a concept was extracted from an actual metadata element, its meanings may not be specific enough to represent a particular characteristic or property of a resource. The concepts derived from superordinate elements functioned as basic facet and subsumed related facets derived from subordinate elements. These related facets divided the broad meaning of the basic facet into more specific units of meaning. Specific facets related to the same basic facet were subsequently categorized based on functional similarities.

Table 4 demonstrates the categorization of facets related to the title of a resource. Although a resource may be represented by one title, in many cases, a resource will have related title elements (e.g., uniform title, abbreviated title, or translated title). It may also have associated values that represent a part of the title, such as a subtitle, part name, or part number. In Table 4, Category 1 identifies the basic facet Title that was extracted from the superordinate elements and represents a single, broad characteristic or property of a resource. The basic facet semantically subsumes the different parts and kinds of titles. Category 2 identifies the facets that describe specific aspects of the basic facet in Category 1 – in this case, the parts (i.e., part-whole relationships) that can be identified in a title. The facets identified in Category 3 are not parts of the broader concept, but represent different forms or kinds of titles (i.e., is-a relationships).

Each category was constructed independently of the other categories based on the functional similarity of facets, but they are closely related: Category 2 represents the parts of the superordinate facet in Category 1; and Category 3 identifies specific types of Category 1. In spite of their different functions, Categories 2 and 3 both serve as containers: That is, each category holds similar yet heterogeneous facets that, taken together, comprise the basic facet. Using these containers, the different functions of each facet and their relationships can be specified and differences in semantic ranges can be mediated (see Figure 1).

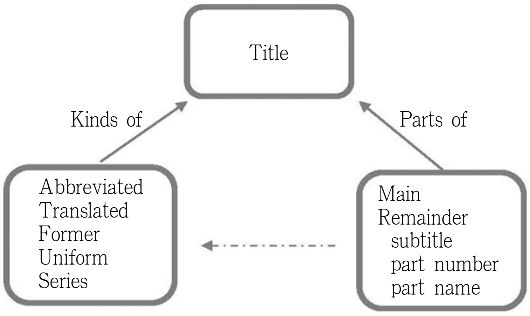
Based on the semantically distinct categories illustrated in Figure 1, a facet structure was constructed

〈Table 4〉 Example of Facet Categorization for the Category Title

Category 1	Category 2	Category 3
Basic facet	Conceptual facet	Relational facet

Title	Main title
	Remainder
	  Subtitle
	  Part number
	  Part name
		Uniform title
		Abbreviated title
		Translated title
		Former title
		Series title

www.kci.go.kr



〈Figure 1〉 Example of Semantically Distinct Relationships for the Title Concept

that categorized constituent facets according to their functional roles. The facet structure represented in Table 5 consists of three components: basic facet, conceptual facet, and relational facet. This set of facets has its own hierarchical structure of relationships between facets: The conceptual facet remainder, which refers to any component of the title except the main title, has two components (subtitle and part). The concepts in the relational facets are not parts of the basic facet Title, but represent different types or kinds of the facet. Importantly, facets in the relational facets can be described using the facets in the conceptual facets because these are more specific

〈Table 5〉 Example of Facet Categorization for the Category Title

Basic Facet	Title
Conceptual Facet (part-of)	main title
	remainder
	subtitle
	part
	part number
	part name
Relational Facet (kind-of)	abbreviated title
	translated title
	alternate title
	uniform title

instances of the basic facets and may have multiple aspects that can be represented using the facets in the Conceptual Facets.

Most of the basic facets include both conceptual and relational facets that indicate part and kind functions of the basic facet. However, in some basic facets, the facets are specific enough to represent the corresponding characteristic of a resource or the characteristic itself is simple enough to be described using only one element.

Although the faceted framework was constructed based on elements extracted from existing standards, they are independent of any particular element in any one standard. The faceted framework is a specification based on the inherent meaning of the metadata element(s) as clarified by semantic relationships between facets. It provides an assemblage of related concepts to guide the combination and application of metadata elements from different standards.

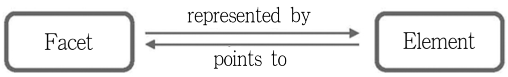
4.5 Implementation of Faceted Framework

Metadata elements are used to describe specific

www.kci.go.kr

aspects of a resource and derive their values from the resource being described. However, the values associated with a resource are not values of the element per se, but of the concept that is the core meaning of the element. In this sense, an element is the representation of a concept. The format of an element can be changed according to the context of the associated standard, as reflected in its structure and syntax, while the concept, which is the translated meaning of the element, is not changed regardless of differing contexts.

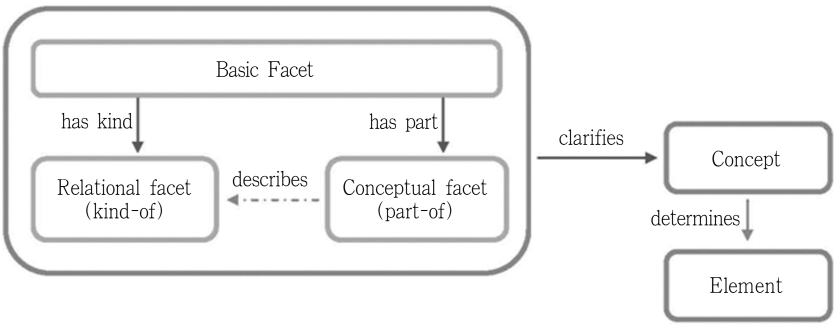
Implementation of the faceted framework is based on the instantiation of facets derived from the vocabulary of concepts that represent the semantic functions of elements. Thus, the faceted framework incorporates three primary components: the elements, the facets, and the functions of the facets. Each of the components is key to implementation of the faceted framework and functions as a node that is connected to other nodes by specific relationships. An Element Node represents a metadata element and prescribes the semantic range of the element based on instantiation of the corresponding facet categories. A Facet Node identifies the core meaning of an element, providing the semantics for the element. The element represents – points to or stands in for – the facet that is its core meaning. The connection between the substantive meaning of the element (the facet) and its representation (the element label) is established through the Facet-Element pair relationship (see Figure 2).



〈Figure 2〉 Facet-Element Pair

The basic facet provides a space that encompasses related conceptual facets and relational facets. The basic facet subsumes the concepts represented by subordinate facets that specify a particular aspect of the basic facet, as represented by the conceptual facet. Different types of the basic facet are specified in the relational facet. These two different facets of the faceted framework are used to clarify similar but ambiguous meanings of concepts based upon the structure and syntax of the standard.

Figure 3 shows the implementation of the faceted



〈Figure 3〉 Implementation of Faceted Framework

www.kci.go.kr

framework with the specification of the Facet Node. The basic facet from superordinate elements functions as a facet category that provides a space to encompass related facets and other components. The basic facet consists of the facets from subordinate elements that functions as a part of and specifies a specific aspect or characteristics of the basic facet (Conceptual Facet Node). Different kinds of the basic facet are specified in the Relational Facet Node identified in the faceted framework. These two nodes of the faceted framework are used to clarify the similar but ambiguous meanings of the concepts extracted from metadata elements by specifying their different functions. Each of 13 basic facets has one framework with subordinate conceptual facet and relational facet based on the facet categorization presented in Table 5 (See Appendix). Although they are constructed independently of other basic facets, they are interrelated and operate together when considering metadata interoperability.

The different types of facets in the Facet Node specify the semantic functions of concepts in Concept Node. The Basic Facet Node prescribes the semantic range of elements. Thus, the concepts included in a certain basic facet can be considered that they describe the same or similar aspect of a resource regardless of the variation of representation. Each Basic Facet Node encompasses conceptual and relational facets based on the relationships between concepts established through the vocabulary of concepts. These subordinate facets categorize interrelated concepts extracted from elements in different standards and remove the semantic conflict resulted from the syntactical and structural differences across standards. This set of facets clarifies the semantics of concepts in Concept Node. This Facet-to-Concept linkage determines the semantics of the elements in Element Node and makes the semantically interoperable.

The faceted framework is not intended to be used for any actual resource, but to provide a set of core elements that can link elements from existing metadata standards. Facet, which represents an element from different standards in context-independent way, can semantically connect the elements in both the MARC and MODS because the facet was extracted from the elements contained in those standards. It was optimized to make metadata elements from different standards interoperable by identifying and linking elements with the same concept and, thus, to establish semantic relationships between those standards. Ideally, the faceted framework generates a conceptual structure that can connect elements from different standards on the basis of semantic, syntactic, and structural similarities by identifying conceptual orientation of seemingly disparate elements within a single, context-independent framework. Simply put, it is the facet-to-concept linkage that makes different standards semantically interoperable through the identification of the semantics of their elements.

The faceted framework is constructed as a conceptual structure that reveals semantic relationships between metadata elements and translates them into the concepts represented in a context-independent framework that can be applied to identify different types of facets. The basic assumption supporting the faceted framework is that elements from different

www.kci.go.kr

standards can be semantically connected if they share a common referent. Semantic connections were established by analyzing the core meaning of each element and identifying common referents through the facet structure in order to establish a set of shared concepts. Subsequent grouping of elements under the same facet prescribed the semantic range and clarified the semantics of each element, while establishing relationships between them. Through the concepts extracted from elements and the different types of facets that categorize the same or similar concepts, elements were linked regardless of different labels and different levels of granularity in the original metadata standards. In addition, a metadata element is understood here as the representation of the concept that is governed by the context reflected in the syntax and structure of the metadata standard in which it appears. Different contexts are mediated in the context-independent faceted framework, and elements from different standards co-exist in the framework based on their semantic similarities. In this way, the faceted framework offers the potential to integrate semantic, syntactic, and structural interoperability within a single framework.

5. Conclusion

The approach to interoperability developed here offers an alternative to existing approaches in its assumption that interoperability can be achieved using a faceted way of representing the core meanings of elements from different standards and linking the semantics of those elements. Through the application of the faceted framework, different standards can be mediated in a context-independent framework that allows metadata elements originating in one context to be interpreted in another context with the clear interpretations of the semantics of each standard. It also provides the ability to combine semantic, syntactic, and structural interoperability within a single framework. With these advantages, the faceted framework provides a context-independent and consistent means for improving metadata interoperability across heterogeneous metadata standards.

This framework is not intended to describe any specific resources but to provide a set of core meanings of metadata elements, that is, facet. This set of facets was categorized according to their semantic similarities through the application of facet analysis. Facet analysis provided three types of facet: basic, conceptual, and relational facet. Basic facet was placed at the top of the structure of the generated data model and shows the framework of the structure. Conceptual facet, placed under each basic facet, defines representative aspects and characteristics of its basic facet. Relational facet specifies different types or kinds of its basic facet. These different types of facets are assigned to each element in both standards based on the semantics of those elements. Through these assigned facets, the elements with the same facets can be connected with each other. This connection makes both MARC and MODS interoperable with each other.

The faceted framework is a conceptual model that provides scaffolding for locating elements from dif-

www.kci.go.kr

ferent standards under the shared (or shareable) concepts. Because it is independent of any context, the faceted framework supports interoperability by demonstrating relationships among elements, concepts, and even values of the elements. Through the faceted framework, metadata interoperability can be achieved not simply at the level of metadata elements, but at the level of semantic concepts. It also provides the capability of integrating semantic and structural interoperability by taking into account the contextual differences between MARC and MODS. Thus, through this faceted framework, interoperability between metadata standards can be more reliable and comprehensive by reflecting the semantic, syntactical, and structural differences across heterogeneous metadata standards.

References

Arms, W. Y., D. Hillmann, C. Lagoze, D. Krafft, R. Marisa, J. Saylor, and C. Terrizzi. 2002. "A spectrum of interoperability: The site for science prototype for the NSDL." *D-Lib Magazine,* 8(1). [cited 2010.5.10]. <http://www.dlib.org/dlib/january02/arms/01arms.html>.

Blanchi, C., and J. Petrone. 2001. "Distributed interoperable metadata registry." *D-Lib Magazine,* 7(12). [cited 2010.5.10]. <http://www.dlib.org/dlib/december01/blanchi/12blanchi.html>.

Chan, L. M. 2005. "Metadata interoperability: A methodological study." *Proceedings of the 3rd China-U.S. Library Conference,* March 23-25, 2005, Shanghai, China, 110-15.

Gill, T. P. and P. Miller. 2002. "Re-inventing the wheel?: Standards, interoperability and digital cultural content." *D-Lib Magazine,* 8(1). [cited 2010.5.10]. <http://www.dlib.org/dlib/january02/gill/01gill.html>.

Jung, H., W. Sung, and D. Park. 2006. "Project report on a Korean Science & Technology Thesaurus with conceptual/relational facets." *Proceedings of 3rd International WordNet Conference,* Seogwipo, South Korea, January 22-26, 2006, 311-313.

Kumar, P.S.G. 1987. *Introduction to Colon Classification*, Dattsons, Nagpur, 1987.

Miller, P. 2000. "Interoperability: What is it and why should I want it?" *Ariadne,* 24. [cited 2010. 5.10]. <http://www.ariadne.ac.uk/issue24/interoperability>.

Moen, W. E. and P. Benardino. 2003. "Assessing metadata utilization: An analysis of MARC content designation use." *Proceedings of International Conference on Dublin Core and Metadata Applications,* DC-2003, September 28 -

www.kci.go.kr

October 2, 2003, Seattle, WA, 171-180.

Oltmans, E. E. 2001. *Metadata Interoperability*. Telematica Institute. [cited 2010.5.10]. <https://doc.telin.nl/dsweb/Get/Document-14690/interoperability.pdf>.

Priss, U. and E. K. Jacob. 1998. “A graphical interface for faceted thesaurus design.” *Proceedings of the 9th ASIS SIG/CR Classification Research Workshop,* American Society for Information Science, Silver Spring, MD, October 25, 1998: 107-118.

Ranganathan, S. R. 1967. *Prolegomena to Library Classification.* New York: Asia Publishing House.

Svenonius, E. 2000. *Intellectual Foundation of Information Organization.* Cambridge, Mass: MIT Press.

Taylor, A.G. 1992. *The Organization of Information.* Englewood, Colo: Libraries Unlimited. <http://www.getty.edu/research/conducting_research/standards/intrometadata/path.html>.

Yang, K., E. K. Jacob, A. Loehrlein, S. Lee, and N. Yu. 2004. “Organizing the Web: Semi-Automatic construction of a faceted scheme.” *Proceedings of IADIS International Conference WWW/Internet 2004,* ICWI 2004, October 6-9, 2004, Madrid, Spain: 374-381.

www.kci.go.kr

Appendix. Facet Categorization for 13 Basic Facets

Basic Facet	Title
Conceptual Facet	main title
	remainder
	subtitle
	part
	part number
	part name
Relational Facet	abbreviated title
	translated title
	alternate title
	uniform title

Basic Facet	Publication
Conceptual Facet	publisher
	name
	address
	place of publication
Relational Facet	--

Basic Facet	Responsibility
Conceptual Facet	name
	family name
	given name
	address
	affiliation
	role
Relational Facet	Personal
	Corporate
	Conference

Basic Facet	Language
Conceptual Facet	--
Relational Facet	--

Basic Facet	Date
Conceptual Facet	issued
	created
	digitized
	modified
	copyright
Relational Facet	Original
	Digital
	Record

Basic Facet	Relation
Conceptual Facet	is version of
	has version
	is replaced by
	replaces
	is required by
	requires
	is part of
	has part
	is referenced by
	references
	is format of
	has format
	source
Relational Facet	--

Basic Facet	Identifier
Conceptual Facet	--
Relational Facet	--

Basic Facet	Resource location
Conceptual Facet	physical location
	digital location
Relational Facet	--

www.kci.go.kr

Basic Facet	Description
Conceptual Facet	abstract
	table of contents
	target audience
	digitization
	technique
Relational Facet	--

Basic Facet	Format
Conceptual Facet	extent
	medium
	physical medium
	digital medium
Relational Facet	--

Basic Facet	Type
Conceptual Facet	--
Relational Facet	--

Basic Facet	Copyright
Conceptual Facet	copyright holder
	access restriction
	copyright date
Relational Facet	--

Basic Facet	Subject
Conceptual Facet	topic
	genre
	classification
	temporal
	geographic
Relational Facet	--

www.kci.go.kr