

Faceted Classification for Museum Artefacts: A Methodology to Support Web Site Development of Large Cultural Organizations

Elaine Ménard, McGill University

elaine.menard@mcgill.ca

Sabine Mas, Université de Montréal

sabine.mas@umontreal.ca

Inge Alberts, Cogniva Information Solutions

i.alberts@cogniva.ca

Abstract

Purpose

The research project aimed to provide a new visual representation of the Artefacts Canada digital collection, as well as a means for users to browse this content. *Artefacts Canada Humanities* is a database containing approximately 3.5 million records describing the different collections of Canadian museums.

Design/methodology/approach

A four-step methodology was adopted for the development of the faceted taxonomy model. First, a *Best Practice Review* consisting in an extensive analysis of existing terminology standards in museum communities and public Web interfaces of large cultural organizations was performed. The second step of the methodology entailed a *Domain Analysis*, consisting in extracting and comparing relevant concepts from terminological authoritative sources. Thirdly, we proceeded to *Term clustering & Entity Listing* which involved the breaking up of the taxonomy domains into potential facets. An *Incremental User Testing* was also realized in order to validate and refine the taxonomy components (facets, values, and relationships).

Findings

The project resulted in a bilingual and expandable vocabulary structure that will further be used to describe the Artefacts Canada database records. The new taxonomy simplifies the representation of complex content by grouping objects into similar facets to classify all records of the Artefacts Canada database. The user-friendly bilingual taxonomy provides worldwide visitors with the mean to better access Canadian virtual museum collections.

Originality/value

Few methodological tools are available for museums who wish to adopt a faceted approach in the development of their Web sites. For practitioners, the methodology developed within this project is a direct contribution to support Web site development of large cultural organizations.

1. Introduction

Nowadays, thousands of museums give access to their collection all over the world through their Web sites. The educational potential of these resources is immeasurable. Unfortunately, these Web sites are often too complex and sophisticated to be useful and usable for the majority of visitors. According to Marty and Twidale (2004), many factors can overwhelm the museum Web site user: overloading amount of multimedia content, very artistically designed graphical user interfaces, interfaces that only encourage exploration but not keyword research, sites designed poorly for museum professionals, and sites oriented to supplement physical museums. Too frequently, the needs of the users of these museum web sites are overlooked in the organization of the museum content. In order to minimize the

aforementioned frustrations and difficulties, faceted classification presents great potential in providing the users with simple, flexible and scalable information access points. However, there are few methodological tools available for museums who wish to adopt such approach in the development of their Web sites.

2. Objective of the project

The research project aimed to provide a visual representation of the *Artefacts Canada* digital collection, as well as a means for the majority of users to browse this content. *Artefacts Canada Humanities* is a database containing approximately 3.5 million records describing the different collections of Canadian museums. This database is at the core of the Canadian Heritage Information Network (CHIN)'s work to make Canada's heritage accessible. CHIN's mission is to connect Canadians and worldwide audiences to Canada's heritage and to promote the development, the presentation and preservation of Canada's digital heritage content for current and future generations.

In order to provide a new visual representation for this digital collection, a faceted taxonomy model was created. The taxonomy creation involved the implementation of an innovative methodology.

3. Methodology

A four-step methodology was adopted for the development of the faceted taxonomy model. First, a *Best Practice Review* consisting in an extensive analysis of existing terminology standards in museum communities and public Web interfaces of large cultural organizations was performed. The second step of the methodology entailed a *Domain Analysis*, consisting in extracting and comparing relevant concepts from terminological authoritative sources. Thirdly, we proceeded to *Term clustering & Entity Listing* which involved the breaking up of the taxonomy domains into potential facets. An *Incremental User Testing* was also realized in order to validate and refine the taxonomy components (facets, values, and relationships).

The four phases of the methodology are described below. The main results are then presented.

3.1 Best Practice Review

The best practices review consisted in an extensive analysis of existing terminology standards in Canadian museum community and public web interfaces for the collection databases of large cultural organizations. The objectives were the following: 1) to acquire knowledge of the industry's terminology standards; 2) to assess how controlled vocabularies can be adapted in the development of the new taxonomy; and 3) to examine Web interfaces for the collection of large cultural organizations to gather important usability criteria prior to the taxonomy's development. Twenty-nine terminology sources (such as the Classification System for Historical Collections from Parks Canada and the McCord Museum Web site) have been analyzed. Since these resources were generally available in English, the English version of the resources was examined even though these were also available in French. The analysis at this stage was more conceptual than terminological.

3.2 Domain Analysis

Since a well defined set of facets can only provide a conceptual representation of one single domain, domain analysis was conducted prior to facet selection. Facet accuracy depends on the subject field homogeneity meaning that a smaller domain leads to more specific and

detailed facets (Mills, 2004). In the case of the museum's collection, the domain corresponds to the perspectives adopted in describing a cultural artefact. For this project, the objective of domain analysis was threefold: 1) to limit the scope and the coverage of the future taxonomy while insuring the taxonomy coverage is accurate and sufficient; 2) to explore and detect potential facets; and 3) to provide recommendations on first-level facet selection. Domain analysis consisted in a mapping of selected vocabulary source categories/fields with Artefacts Canada database fields and an analysis of Artefacts Canada database's frequency reports. The analysis of the Artefacts Canada frequency reports aimed at selecting candidate facets by identifying important concepts represented in the various fields of the actual database. The proposed strategy for the analysis of the frequency reports, in respect of the twenty-four fields of Artefacts Canada database, enclosed two levels of analysis: a conceptual analysis (i.e. recognition of the most important concepts represented by the terms and detection of terminological issues within the lists of terms), and a statistical analysis (i.e. selection of candidate terms and selection of terms to be removed following their statistical value).

3.3 Term clustering and selection of the top-level facets and sub-facets

The third phase in the process of the taxonomy model development involved the choices of top-level facets and their sub-facets. Two parallel approaches were combined: a bottom-up approach and a top-down approach. Starting from a list of potential facets (such as "Artist", "Date" or "Style") resulting from domain analysis, the relevant values were extracted from the Artefacts Canada frequency reports and these values were clustered into a limited number of classes. These classes were the basis from which facets, sub-facets, and candidate descriptors were selected. For the taxonomy development, frequencies of the terms in the different Artefacts Canada fields were used as a form of weighting to assess a list of candidate terms. During the frequency report analysis, we realized that a saturation point was reached beyond the first 500 single terms whose frequency of use was the highest. Beyond that point, a great majority of terms were becoming repetitive. A more conceptual analysis of Canadian museum collections was also carried out, inspired from the best practice review. The vocabulary sources analyzed previously were used for the selection and the definition of the top-level facets and their sub-facets. The facet selection and taxonomy development phase was incremental, implying the combination of a bottom-up approach with a top-down strategy at different stages of the project.

3.4 Incremental User Testing

During the taxonomy development, a two-phased user testing strategy was conducted to insure the final product was clear, comprehensive, and consistent. In total, eighteen different users contributed to the development of the taxonomy.

The card sorting technique was employed during the first evaluation. This data collection method is mainly used in fields such as Psychology, Cognitive Science, and Web Usability. The card sorting technique consists in a classification exercise where participants are asked to organize a set of "cards" presenting different categories. The card sorting technique allows the confirmation of a given information structure and the understandability of the proposed labels.

The user sample included 5 French speaking persons (1 man, 4 women) and 3 English speaking persons (3 men) from different age groups. Sixty cards matching the sixty second level labels of the taxonomy draft were presented to each participant. Participants were asked to proceed to a card sorting exercise, i.e. organize the set of cards by groups of information. Once

all the sorting process was completed, the participants were asked to put a note on the top of each pile and write a title for each specific category they formed. Various questions were asked to the participants in order to understand the rationale of their sorting, in comparison with the taxonomy draft. This step was especially useful to understand the reason why the cards of a group were classified together. Globally, the card sorting exercise led to the validation of the first draft of the taxonomy, specifically its accuracy with regard to the mental representations that potential users might have when accessing the Artefacts Canada collection.

The second phase of user testing entailed an evaluation of the performance of the taxonomy usage, involving a usability test under experimental conditions. First, a simulation of the retrieval process was performed. The objective of this testing was to ask a representative sample of users to complete typical retrieval tasks using the proposed taxonomy in order to measure the degree of effectiveness, efficiency, and user satisfaction. The performance testing was expected to identify usability inconveniences of the new taxonomy that may not be revealed by less formal testing. The experiment also aimed to evaluate the quality of the vocabulary, the structure of the taxonomy, and the selection of specific facets.

Ten participants from different age groups and various experiences in museum collections (primary school students, regular workers, retired people, etc.) accepted to participate on a volunteer basis. Six of the participants were English native speakers and four were French native speakers. Using a mock-up of the taxonomy structure, participants were asked to retrieve fifteen museum artefacts. During the retrieval process, the following variables were recorded: number of choices used to complete the retrieval task, facets selected by the participants for each query (object to retrieve) and additional comment on the taxonomy. For the retrieval simulation, fifteen objects were randomly selected from the Artefacts Canada database. These objects were searched in the same order by all the participants using the taxonomy version in their own language. Once the retrieval simulation was completed, a written questionnaire was submitted to the participants in order to gain their general opinions and to report any difficulties encountered during the retrieval process. The questionnaire evaluated the quality of the entire taxonomy as well as the overall satisfaction from an end-user perspective.

4. Results

The first part of the best practices analysis entailed the review of terminology standards and background documentation used in the Canadian museum community. We examined 13 relevant resources. This review was based on two specific dimensions: the coverage, concepts, and categories used to describe museum collections, and the general best practices and applications used for a specific museum collection. From our analysis, we observed that the majority of the reviewed resources covered many different aspects of the museum's collections. The depth and complexity of these aspects vary significantly from one resource to another, depending on many factors such as local policies and user needs. According to the first phase of our analysis we can divide the elements used to describe the museum's collections in two broad categories: 1) *descriptive elements*, i.e. elements which describe the content of the object and 2) *contextual elements*, i.e. elements which describe the context and the form of the object. Our analysis also brought to attention that the 13 examined resources present different approaches to organize information. In fact, the majority of the examined resources provide a hierarchical structure, i.e. using first-level elements and in most cases, second-level elements, third-level elements, and so on. According to the examined resources, there is no standard number of hierarchical levels. Moreover, the hierarchical structures are in many cases very deep while very limited in others. Consequently, we did not distinguish a particular tendency in the organization

of the second level of the identified first level elements. When examining the thirteen mentioned resources, we observed that six resources exist only in English while five of them are bilingual. Consequently, an individual having little knowledge of the English language and wishing to use these resources will face a major linguistic problem unless a translation mechanism is offered to facilitate their use. Since the target audience of the future taxonomy is bilingual and most resources used by the museum only exist in English, it was essential to consider the linguistic aspect in the development of the taxonomy.

For the second part of the best practices analysis, 16 selected public Web interfaces designed for the collections databases of large cultural organizations were examined. This analysis was mostly based on standards used to describe online collections yet also considered the many aspects of usability including search and browse functionalities offered by these Web sites. The examination of the 16 Web sites revealed a great variability in the organization of the information. The majority of the sites have a tendency to organize the information in a hierarchical structure, mostly with a first level structure (including 2 to 9 first-level facets) and second level elements more or less developed according to the collections offered online. However, the examination of these Web sites did not determine a particular tendency in the organization of the second level of the identified first level elements. Finally, we noticed that some Web sites organize the access to their collections according to the targeted public (young children, students, etc). The Web site is therefore a medium for information sharing and offers a platform for representing and promoting the collection of the museum. Interface and information usability refers to the ease in which the Web site can be used by its intended audience to achieve defined goals and must take into account factors such as information structure, design, and search functionalities to ensure the best access possible.

The examination of the 16 mentioned Web sites again revealed disparities from one museum to another, particularly in how information is organized, interfaced, and the existing search and browse functionalities. Consequently, the users of several Web sites can quickly be overwhelmed by Web sites which are especially dedicated to museum professionals (terminology or categorization of the information being too difficult to understand for a non professional) or by Web sites offering an interface that requires an extended learning process. Finally, the majority of the Web sites examined (12 of 16) only offer English interface which greatly limits the target audience. In brief, the analysis of the best practices provided recommendations regarding the development of the Artefacts Canada taxonomy. The recommendations focused on the top-level facet definition and their organization for the future taxonomy users.

Domain analysis consisted in comparing the fields used in the Artefacts Canada database with those used in the resources analyzed during the best practices review. The conceptual and statistical analysis of the frequency reports first allowed: to carry out a categorization of the fields, i.e. to classify the fields themselves into five categories: 1) classifying fields; 2) fields relating to the description and the composition of the object; 3) fields relating to time; 4) fields relating to the place of creation and use of the object; and 5) fields relating to the artistic, cultural and academic context in which the object was created. It also allowed considering several conceptual regroupings and subdivisions. Finally, the domain analysis led to the identification of the existence of a relation between several fields and the relevant fields for later facet identification. Resulting from domain analysis, we were able to limit the scope and the coverage of the future taxonomy while insuring the taxonomy coverage would be accurate and sufficient, explore and detect potential facets and provide general recommendations on first-level facet selection.

Three essential steps were followed in the development of the classification structure. Firstly, the taxonomy domain was broken up into elemental concepts. These concepts were gathered into fundamental classes. These classes were the basis from which facets were selected. Secondly, the facets were organized into a coherent structure realized simultaneously with the choice and definition of top-level facets and their sub-facets. The number of top-level facets, as well as the depth of the taxonomic structure, was restricted. The number of first level facets was limited to seven and the level of depth has been reduced to four, developing up to five in some instances. Thirdly, the last stage of the facet and sub-facet selection process entailed the normalization of the facets and their descriptors using authoritative sources. The facets and their descriptors were first developed in English while keeping track of French resources when available.

The results of the user testing are twofold. Card sorting results were compiled and analyzed in order to compare the piles created by the users and the naming assigned to each of the pile. According to the results obtained from the card sorting exercise and from user comments, several changes were brought to the first draft of the taxonomy. For example, the facet "Subject" was renamed "Subject Represented", the facet "Time" was merged with "Style & Period" and the discipline facet was "flagged" as a difficult facet for users, requiring further validation during the second phase of the user validation process (See Figure 1).

The second phase of user testing revealed that users mostly focus on facets such as "Geographical Location", "Style & Period" and "Object Type" during the retrieval simulation. But this does not indicate that the other facets were not efficient, but rather that these choices were mostly driven by the queries designed to generate the browsing. Due to time limitation, it was not possible to explore all the possibilities offered by the taxonomy. However these results highlighted the value of the new browsing interface. The number of facets used by the participants for multiple-concept queries was also an interesting outcome of the experimentation. The results obtained indicated that users are willing to search objects through a combination of categories. In other words, the user testing emphasized the enormous potential of facet classification and the endless possibilities offered to the users to browse the database with multiple entry points. The user testing also demonstrated that the approach in designing the new taxonomy answered the fundamental questions of *who* is doing something, *what* is being done, *how* it is being done, *by what means*, *where* and *when* it is being done. Each first level facet is a specific entity that can be employed individually as well as in combination. Finally, we were particularly encouraged by the response of the participants stating the structure easiness of use and confirming their enthusiasm to browse the Artefacts Canada database with this taxonomy.

The final version of Artefacts Canada taxonomy includes seven facets:

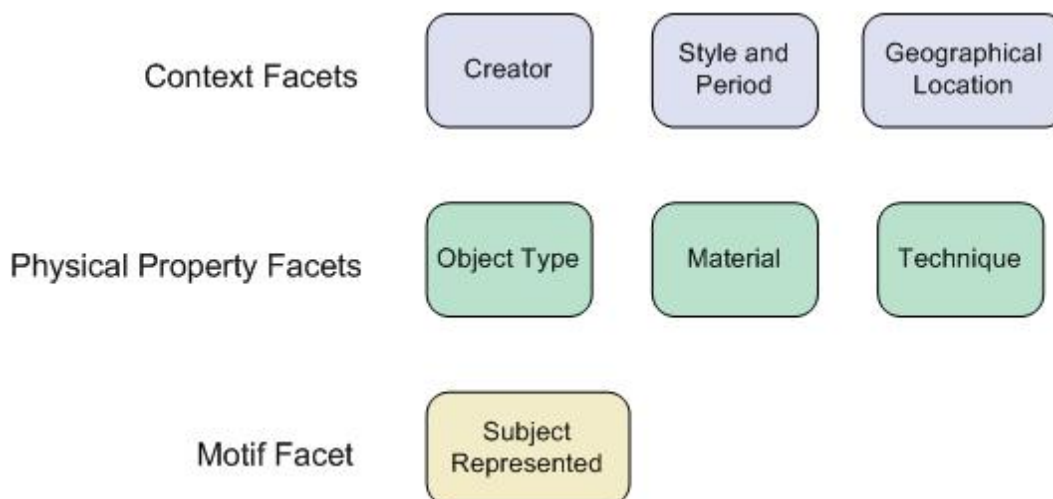


Figure 1. Overview of Artefacts Canada Taxonomy Model

As depicted in figure 1, three main perspectives have oriented the choice of the top-level facets:

1. **Context Facets:** Facets relating to the set of facts or circumstances surrounding the creation of the artefact;
2. **Physical Property Facets:** Facets that constitute properties characterizing the apparent artefact's nature, form or function;
3. **Motif Facet:** Facet representing the unifying idea being communicated or rendered by the artefact.

These perspectives constitute the rationale behind the designation of the seven top-level facets aiming to describe an artefact from its essential angles. Each of the seven facets is briefly defined in table 1.

Facet	Definition	Sub-Facets
Creator	Definition: The <i>Creator</i> facet refers to the entity that creates or produces the object or the work. Goal: This facet aims to give access to the different entities that produce the objects of the database.	Artist/Maker Manufacturer
Geographical Location	Definition: The <i>Geographical Location</i> facet contains the name of the general regions, continents, countries and provinces where the object was created (for work of art) or use (for functional object). Within this facet, cultural information is also presented, using subdivisions such as "historical region" or "former nation". Goal: This facet aims to provide an access to different types of objects following their geographical origin of creation or usage.	Canada Continent Former Nation, State or Empire Historical Region World Region
Material	Definition: The <i>Material</i> facet refers to what the artefact is comprised of, made with. Most of the terms regrouped in this facet represent material based on their composition or origin. Goal: This facets aims to group objects following their composition or substance	Animal Material Clay Fibre and Textile Metal Pigment, Paint and finish Plant Material Pulp, Paper and Derived Products Stone and Mineral Synthetic Material
Object Type	Definition: The <i>Object Type</i> facet refers to all entities	Art Object

Facet	Definition	Sub-Facets
	<p>which can be perceived by the senses, and most specifically by the view.</p> <p>Goal: This facet aims to provide an easy access to different types of objects recorded in the database, from works of art to common objects.</p>	Functional Object
Style and Period	<p>Definition: The <i>Style and Period</i> facet provides the names of distinct historical periods, broad cultural region styles and periods, art and architecture movements and groups & schools that are represented in the work being classified.</p> <p>Goal: This facet aims to provide an access to different types of works recorded in the database, from a historical, cultural or stylistic point of view.</p>	Movement, Group and School Period Regional Style and Period
Subject Represented	<p>Definition: The <i>Subject Represented</i> facet refers to the iconographic content to describe the subject or image of the object whether the optical counterpart exists in reality or in imagination.</p> <p>Goal: This facet aims to provide an access to different types of subjects matter represented in the database like a person, an animal, a landscape, an object, a building, a place, or an event.</p>	Animal Representation Architectural Representation Event Representation Human Representation Nature Representation Object Representation Place Representation
Technique	<p>Definition: The <i>Technique</i> facet represents the processes, methods and means used to produce an object. This facet contains techniques used to produce objects according to their materials as well as more specific techniques used for creating art objects.</p> <p>Goal: This facet aims to provide an access to different types of objects following the technique used during the production process.</p>	Art Technique Material-Working Technique

Table 1. Facets Definition and Goal

5. Discussion and conclusion

Following this analysis of museum resources and Web sites, we recommend that facet organization should be based on: facilitating the learning process, helping users remembering how to use the Web site, reducing the likelihood of errors, and increasing efficiency and effectiveness in order to encounter a higher degree of user satisfaction. More specifically, we also recommend that the general organization of taxonomies should take into account the users' perspective in order to increase the level of satisfaction. To that end, taxonomy developers should limit the number of choices presented to the users in order to avoid frustration and increase the degree of satisfaction. Finally, we suggest that the interface, searching, and browsing functionalities in French and English should be as simple as possible in order to improve the efficiency and the effectiveness of the user searches.

In the 1970s and 1980s, many cognitive scientists studied the way in which people naturally classify and categorize things. Rosch (1978) defined two principles of categorization. The first is "cognitive economy", which states that classifications exist in order to provide the greatest amount of information with the least amount of cognitive effort. The second principle is the "perceived world structure", i.e. the world as we perceive it gets delivered to us as structured information, not in an arbitrary manner. Following the analysis of several resources and Web sites, we recommend organizing the information in a manner which offers a cognitive economy in addition to a definite hierarchical structure. For that aim, the taxonomy resulting from this

project includes essential facets, both in English and in French, which appropriately represent the resource in order to increase the effectiveness and the efficiency of the user browsing and searching; facets which present the appropriate level of granularity in order to improve the effectiveness and the efficiency of the user searches; and finally, clear and available definitions in order to enhance the satisfaction of the taxonomy by users.

The resulting bilingual and expandable vocabulary structure will further be used to describe the Artefacts Canada database records. Three main perspectives have oriented the choice of the top-level facets: the context facets, i.e. facets relating to the set of facts or circumstances surrounding the creation of the artefact; the physical property facets, i.e. the facets that constitute properties characterizing the apparent artefact's nature, form or function; and the motif facet, i.e. the facet representing the unifying idea being communicated or rendered by the artefact. These perspectives constitute the rationale behind the designation of the seven top-level facets aiming to describe an artefact from its essential angles. This structure offers a high-level and flexible representation of the information accessible in the database. The new taxonomy also helps representing the complexity of the content and to group objects into similar concepts to classify all records of the Artefacts Canada database. The user-friendly bilingual taxonomy will provide worldwide visitors with the mean to better access Canadian virtual museum collections. For practitioners, the methodology developed within this project is a direct contribution to support Web site development of large cultural organizations.

References

- Marty, P. F., and Twidale, M. B. (2004), "Lost in gallery space: A conceptual framework for analyzing the usability flaws of museum Web sites", *First Monday*, Vol. 9 No 9, http://firstmonday.org/issues/issue9_9/marty/index.html (accessed 22 October 2008).
- Mills, J. (2004), "Faceted Classification and Logical Division in Information Retrieval", *Library Trends*, Vol. 52 No. 3, pp. 54-570.
- Rosch, E. (1978), "Principles of categorization", in A. Collins & Smith E. E. (Eds.), *Readings in cognitive science, a perspective from psychology and artificial intelligence*, Morgan Kaufmann Publishers, San Mateo, CA, pp. 312-322.