Carlos Henrique Marcondes – Federal University of Minas, Brazil
Célia da Consolação Dias – Federal University of Minas Gerais, Brazil

Representing Faceted Classification in SKOS

Abstract:
Faceted classification is one of the most important contributions from knowledge organization to information resources management. Proposed in the twentieth century by pioneers such as Bliss and Ranganathan as an alternative to the rigidity of enumerative classification, faceted classification has gained importance, both academic and practical. Today it is widely used in the web architecture, from scientific to commercial sites, and even as a methodology for the development of ontologies. The web is evolving towards the vision of the Semantic Web, in which resources are not only navigable and understandable by humans, but also with content that has precise meaning. This feature enables computer applications to perform sophisticated tasks. A key issue with assigning meaning to the descriptions of web resources is the vocabularies. Many such vocabularies are faceted. Can faceted classification play a role in the Semantic Web? How could SKOS be extended to represent faceted classification? What type of entity is a facet? What are its components? Could faceted classification be formalized to be represented in Semantic Web standards? Could the current KOS evolve to take advantage of the potentialities of Semantic Web technologies? The aim of this paper is to achieve a conceptual model of a faceted classification and its components; to code this model in SKOS and to evaluate such codification. Canonical definitions of “facet” and its components are used to develop a semantic model of a facet schema and its components. Based on this model, a proposal of codification in SKOS is achieved and evaluated.

1.0 Introduction
In her article “A Semantic (Faced) Web?” La Barre asks the question of how to integrate faceted classification into the Semantic Web:

“The chief focus is upon Semantic Web implementations that employ, adapt, or misconstrue the theory or practice of facet analysis and Faceted Classification. A secondary focus is upon suggestions for the creation of operational definitions and functional requirements for facet theory that may serve to enhance, amplify or extend current understandings and practices in Semantic Web implementations.” (La Barre 2010, 103).

Faceted classification is one of the most important contributions from the knowledge organization to information resources management. Proposed in the twentieth century by pioneers such as Bliss and Ranganathan as an alternative to the rigidity of enumerative classification, faceted classification has gained importance, both academic and practical. Today it is widely used in web architecture, from scientific to commercial sites (Vickery 2008; La Barre 2006, 50), as an information retrieval device (Broughton 2006) and even as a methodology for the development of ontologies (Prieto-Diaz 2002).

Hudon (2019) reminds many authors of the importance of faceted classification applications, such as serving as navigational tool for websites, structuring systems of objects and information about them, and assisting in the understanding of the complex relationships between objects. Broughton and Slavic (2007, 728) stress that “...the potential for faceted approaches to information retrieval in electronic environments had been perceived as early as the beginning of the 1980.”

The web is evolving towards the vision of the Semantic Web, in which resources are not only navigable and understandable by humans, but also by machines. Semantic Web applications navigating between resources can process such resources to perform sophisticated tasks. Within this context, interoperability is key (Zeng 2019), so that generic Semantic Web applications can interact with web resources. Meaning is assign
to such web resources by describing them with different vocabularies. To enable Sem-
antic Web applications to interact with web resources, they must be described for-
mally and with standards languages whose constructs make reference through In-
ternationalized Resource Identifier (IRI)\(^1\) to vocabularies where terms have precise
meaning and global scope. Recently several vocabularies developed for information
retrieval, library systems or databases have been adapted for Semantic Web technolo-
gies and for reference through IRI; many of them incorporate facets. To provide a
bridge between Knowledge Organization System (KOS) and the Semantic Web, a
standard, Simple Knowledge Organization System (SKOS), a metadata model to de-
scribe web resources in the Resource Description Framework (RDF)\(^2\), has been under
development; since 2009 SKOS is a W3C standard. However, the present version of
SKOS vocabulary does not support the representation of faceted KOS (La Barre 2006,
116) (W3C 2009).

How could SKOS be extended to represent faceted classification? What type of en-
tity is a facet? What are its components? Could faceted classification be formalized in
Semantic Web standards (La Barre 2006, 111; Miles and Bechhofer 2008) to take ad-
vantage of the potential of such technologies?

The aim of this paper is to represent a faceted schema in SKOS; in order to achieve
it, a conceptual model of a faceted classification and its components was developed;
such a model was coded in SKOS, and such codification was evaluated.

2.0 The method

Canonical definitions of facets and their components found in KOS literature are
used as bases to identify components of a facet. Sources discussing and defining meta-
physical entities as classes, subclasses, instances, properties, attributes, characteristics
and relationships are used for achieving a semantic model of a faceted schema. Based
on this model a proposal for codification in SKOS is developed and evaluated. Te-
maTres (https://www.vocabularyserver.com/) software is used to generate the cod-
ification in SKOS.

3.0 Results

This section contains an analysis of definitions of the concept “facet” to use as bases
to develop a conceptual model of a facet schema and its components. On the basis of
such a model, a codification in SKOS is achieved.

3.1 What is a facet? What is faceted classification?

La Barre (2003) observed that there is no consensus from scholars for the meaning of
term “facet”.

To understanding the meaning of “facet” canonical definitions were selected from
KOS literature. While Ranganathan (1967a) emphasizes the aspects of a basic subject
and its compound subjects, Mills and Broughton (1977) focus on subclasses and their
principles of division, Soergel (1995) sees facets as entities and finally Taylor (1992)

---

\(^1\) IRI - Internationalized Resource Identifier, https://en.wikipedia.org/wiki/Internationalized_Resource_Identifier

\(^2\) RDF, The Resource Description Framework, is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model
discusses a vision of classes specifically concerning their aspects, properties, and characteristics.

Proposed in the twentieth century by Ranganathan and others faceted classification is an alternative to the rigidity of enumerative bibliographic classifications in which a book is assigned to one class its general subject. Such classifications are problematic when a book has more than one subject or “points of view” or facets, such as the book “Control of virus diseases of the stem of rice plant in the winter of 1967 in Madras” (Ranganathan 1967b, 13). In faceted classification, each subject within a compound subject is considered a facet in generating by a synthetic process, a notation used to locate the book in a specific shelf and retrieve it. Thus faceted classifications are also information retrieval devices used to retrieve entities’ representations based on their properties – where those are the different subjects of complex subjects as suggested by Ranganathan or product’s characteristics in an e-commerce site.

Ranganathan (1967b, 88) conceived facet as “a generic term used to denote any component - be it a basic subject or an isolate - of a compound Subject, and also its respective ranked forms, terms and numbers.”

Soergel (1995, 258) says, “facets are aspects or viewpoints from which entities - such as food products or subjects (topics, themes) in an area such as education - can be analysed.”

Svenonius (2000, 139) sees facets as categories of generality, defining as “grouping of terms obtained by the first division of a subject discipline into homogeneous or semantically cohesive categories.”

Facets are defined by Taylor (1992, 274) as “clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject” or, by Mills and Broughton (1977, 38), who write "A facet may be defined as the total set of subclasses produced when a class is divided by a single broad principle..." Both definitions introduce the concept of class and its division based on its aspects, properties, or characteristics.

A classification schema is defined as “a list of classes arranged according to a set of pre-defined principles for the purpose of organizing items in a collection or entries in an index, bibliography or catalog into groups based on their similarities and differences to facilitate access and retrieval.” (Fallucchi and De Luca 2018). Another definition of schema classification highlighted by Jacob (2004, 524) agrees that “a classification scheme is a set of mutually exclusive and nonoverlapping classes arranged within a hierarchical structure and reflecting a predetermined ordering of reality.” A faceted schema is also a structure to represent entities and relationships.

According to NISO (2005) faceted analysis is a way of organizing knowledge. Facet analysis is particularly useful for:

• new and emerging fields where there is incomplete domain knowledge or where relationships between the content objects are unknown or poorly defined;
• interdisciplinary areas where there is more than one perspective on how to look at a content object or where combinations of concepts are needed;
• vocabularies where multiple hierarchies are required but can be inadequate due to difficulty in defining their clear boundaries; or
• classifying electronic documents and content objects where location and collocation of materials is not an important issue.” NISO (2005, 13).
Although Ranganathan conceives facets within the scope of the bibliographic classifications of books, contrary to enumerative bibliographic classifications in which to a book is assigned only one subject that defines its position within the classification schema, Ranganathan realized that books could be about several subjects simultaneously. Those component subjects of a compound subject, as previously mentioned, are its facets. A subject is a relationship between a book and what it is about; a component of a subject is a component of any (ontologically) thing a book is about. Conceptual models such as FRBR (IFLA 1997) and LRM (Riva et al 2017) consider subject according to the same ontological view.

3.2 What are the components of a faceted schema?

According to De Grolier (1965, 102), “the term facet itself is just a new, fashionable, word for designating the series of subdivisions of a given subject according to one, and one only, of its characteristics.”

Facet analysis aims to meet users’ needs to access information: “what entities [and] what aspects of those entities are of interest to the user group” throughout the process of conceptual analysis” (Vickery 1960, 11). This is a fundamental criterion in the development of faceted classification.

In an example cited in the ISO 25964-1 (2011, 69) standard, there is a subclass of the agricultural industries products class, dairy products, milk: properties of milk, as milk by fat content, milk by form, milk by source animal; and milk by treatment type. From the milk by fat content property, the following facets are derived, giving rise to subclasses: whole milk, low-fat milk, and skim milk. From the property milk by form are derived the following facets or subclasses: dried milk and liquid milk. From the property milk by source animal the following facets or subclasses are derived: buffalo milk, cow milk, goat milk, and sheep milk; and from the property milk by treatment are derived the following facets or subclasses: condensed milk, evaporated milk, homogenized milk, pasteurized milk, and sterilized milk. To this schema, we added, to help develop our arguments, another class, producer, with three subclasses (two of them milk producers and other a tractor producers), Nestle Massey Ferguson, and Parmalat.

We can distinguish the following elements in this example: the class, milk, one of its subclasses according to a criterion or facet (a milk property, the treatment applied to the milk), milk by treatment; and the subclasses derived from applying this criterion, condensed milk, evaporated milk, homogenized milk, pasteurized milk, and sterilized milk.

Vickery (2008, 156), discussing the structure of a faceted classification, distinguishes elements such as D, a Domain, S1, S2, S3..., Subject fields within a Domain, F1, F2, F3... facets within each Subject field, T1, T2, T3... terms within each facet, and the order of terms within each facet. Within the elements listed by Vickery, Subject fields may be associated with Classes in Taylor’s (1992, 274) previous definition.

Three elements can be identified in these definitions: a class; the set of subclasses, generated based on its aspects; and properties or characteristics.

In the previous examples and in many others, a facet is always defined relative to a class by applying to it a criterion, there is not a facet that is not a facet of a class.
According to Ranganathan (1967b, 55), “Characteristic – an attribute or any attribute complex with reference to which the like or unlikeness of entities can be determined and at least two of them are unlike.”

Facets might be those entities identified in metaphysics and ontological analysis as “properties.” Properties are existentially dependent entities, as their existence depends on the existence of the entities that are the bearers of these properties; a specific marriage cannot exist without the existence of the individuals forming the couple. A specific blue colour cannot exist without being the colour of a specific blue object, such as my blue shoes. Marriage and colour are specific types of properties; the former is a relationship, the last is an attribute (Guarino 1997).

Web languages such as Ontology Web Language (OWL) and the Resource Description Framework Scheme (RDFS) distinguish class-subclass properties, or unary properties, from binary properties. The first type comprises the taxonomic or paradigmatic structure of a domain. The two former types of properties are called in OWL an Object Property (a relationship) and Data Property (an attribute). Object properties require the existence of two entities, one being the domain and the other the range of a specific relationship. A marriage is a typical Object Property. Data Property requires just one entity, the domain of a Data Property, while the range is a set of possible values. A colour is a typical Data Property, as it assigns a data value, “blue” to the property colour of an entity. Sowa (2000, 32) makes same distinction between attributes and characteristics.

“Properties (also called ‘attributes,’ ‘qualities,’ ‘features,’ ‘characteristics,’ ‘types’) are those entities that can be predicated of things or, in other words, attributed to them” (Orilia and Swoyer 2011). To Aristotle (1991, 3) the notion of a property, or characteristic, rests on that of predicates; properties are predicated of subjects; hence they do not exist without being properties of something “(for all colour is in a body” as they are dependent of something to which they are predicated.

The same notion is also in Chen’s Entity-Relationship Model (1976): a domain can be modelled identifying the entities, the relationships and the attributes of entities and relationships.

“A substance— that which is called a substance most strictly, primarily, and most of all—is that which is neither said of a subject nor in a subject, e.g. the individual man or the individual horse. The species in which the things primarily called substances are, are called secondary substances, as also are the genera of these species” (Aristotle 1991, 4).

In this citation, Aristotle relates (first) substance to secondary substances, the genera and the species, defining a class-subclass relationship. A (secondary) substance holds the essence of an entity. Substances are organized in hierarchies of class-subclasses in which subclasses are defined by their essences as having the genus of its parent class plus a differentia from it, in a sequence of increasing specificity. Aristotle (1991, 44) also distinguishes “an accident or property of a thing,” those categories that qualify a subject: quantities, qualities, relations, location, time: “Of things said without any combination, each signifies either substance or quantity or qualification or a relative or where or when or being-in-a-position or having or doing or being-affected” (Aristotle 1991, 3).

Applying a property to a class generates subclasses (classes and subclasses are universals, secondary substances according to Aristotle), the individuals or instances that
make up the extension (Orilia and Swoyer 2011, section 1.1.4) of the class (first substances according to Aristotle).

The property (also a secondary substance) is a subclass of, or is a type of, relates a class and a subclass. The property is an instance of relates a class and its instances. Frické (2010) call this kind of instantiation “first-order instantiation,” or the first-order property. He claims that there is a second kind, the property is an instance of, as in the case of Aristotle citation, the second-order instantiation or the application of second-order properties to first-order properties; that is the case of ‘being a species’ applied to the first order property ‘being a tiger’ – the genera. Tiger is an instance of the second-order type species. According to this perspective, organizing books on poetry, prose, or theatre by literary genre means applying a secondary property (the literary genus) to primary properties (poetry, prose, theatre), the values of literary genus assigned to each book. The concept of second-order property is similar to that of meta-property (Guarino 1997).

Applying Frické conceptualization to the ISO 25964-1 example:
- Shere Khan (first-order instance) -> tiger (genus) -> animal (species);
- condensed milk, etc. (second-order instance, subclass) -> milk by treatment (property) -> milk (class).

A property (or characteristic or facet) divides a class, generating different subclasses, one for each different value of that property existing within the domain. According a facet is a non-sortal or characterizing property of a class (Orilia and Swoyer 2011, section 7.8), generating subclasses but no individuals. Faceted classification is not concerned with the class-subclass hierarchy, a “classification ontology” according to Giunchiglia, Dutta and Maltese (2014, 52) within a domain but rather in a given a class, finding the properties of this class, a “descriptive ontology” of interest to users (Giunchiglia, Dutta, and Maltese 2014, 53), and the instances within the domain with different values to each to these properties. The same notion exists in knowledge organization, between paradigmatic relationships, those permanent, structural or taxonomic relations within a domain, and syntagmatic relationships, those ad hoc, a posteriori or transient relationships (Khoo and Na 2006, 164).

First-order logic and languages such as OWL do not deal well with second order properties as the scope of quantifiers range over individuals (Väänänen 2019). However second order properties are appropriate to specify faceted classification, as they raise not only the individuals that are instances of a class but also the subclasses that are instances of a class, according to Frické (2010).

3.3 How to map components into a conceptual model?

Faceted classification have as components classes, their facets, i.e., their properties (unary properties – subclasses -, and binary properties - relationships), which constitute the criteria to derive instances, and the instances themselves (first- and second-order instances) of each specific facet. Eventually, the order of instances within each facet is also specified. There is also a difference between the two types of facets. One is those derived from classes-subclasses relationships, where the two relata are subclasses of just one primary class within the domain, while the other is those derived from relationships in which the two relata belong, or are subclasses of, different primary classes within the domain.
According to Frické (2010, 44) first-order logic is adequate to discuss second order properties. Applying this conceptualization proposed by the author, plus the OWL concept of properties (Data and Object Properties) of each having a domain and a range. A logic theory is proposed in order to formalize the results of the previous analysis.

- Be D1 a domain formed by the primary classes (or unary properties) C1(x1), C2(x2), C3(x3) …;
- Be C1(x1), one such classes;
- Be c1 = \{x1: C1(x1)\} (c1 is the concept of C1);
- Be P1, P2, P3… the properties of the class C1(x1);
- Be P1 = \{a1, a2, a3, an …\} (a1, a2, a3, an … are the extension of P1, i.e., instances of P1).
- Be p1 = \{x1, y1: P1(x1, y1)\} (p1 is the concept of P1);
- Be P1(c1, p1) i.e., the binary properties with domain c1;
- Definition: F1C1(c1, P1): P1(c1, p1) (Facet F1C1 is the class defined by the predicate, or criterion P1, having as domain the concept of class c1 (\{c1\}) and as range the instances of P1, i.e. the class defined by instances of the relationship P1, as for example, in the case of P1 = milk by treatment, the set s1: \{condensed milk, evaporated milk, homogenized milk, pasteurized milk, and sterilized milk \}. In this example, as milk by treatment is one of the attributes of milk, FC1 maybe a Data Property facet in OWL sense.
- Definition: FC1(c1, p): \forall R(c1, p) (the Facets of a class C1 are all the binaries properties that have as domain the class c1).

Another example, be P2 the property milk by producer. F2C2(c1, P2) is the Facet of C1, i.e. the class formed by the set s2 \{Nestle, Parmalat\}. In this example, as producer is a different class from milk, FC2 may be an Object Property facet in the OWL sense.
- Be FC1 = \{a1, a2, a3, an …\} (a1, a2, a3, an … are the instances of P1).
- Be FC2 = \{b1, b2, b3, bn …\} (b1, b2, b3, bn … are the instances of P2).
- Definition: FC1 → □ (an ≠ am) (two second order instances generated by the same facet cannot be equal).
- Definition: F1C1, F2C2 → ◊ (a = b) (it is possible that two second-order instances generated by different facets be equal).

### 3.4 How to map the conceptual model into SKOS?

This section presents the mapping of the conceptual model’s examples of facets in SKOS. Due to page limitations on the number not all the concepts in the example of ISO 25964-1 are presented, only those needed to illustrate the proposal.

```xml
<skos:Concept rdf:about="http://localhost/tematres3.0/vocab/?tema=16">
  <skos:prefLabel xml:lang="en">milk</skos:prefLabel>
  <skos:inScheme rdf:resource="http://localhost/tematres3.0/vocab/>
  <skos:broader rdf:resource="http://localhost/tematres3.0/vocab/?tema=12"/>
  <skos:hasfacet rdf:resource="http://localhost/tematres3.0/vocab/?tema=24"/>
  (source_animal)
  <skos:narrower rdf:resource="http://localhost/tematres3.0/vocab/?tema=17"/>
  <skos:narrower rdf:resource="http://localhost/tematres3.0/vocab/?tema=21"/>
  <skos:narrower rdf:resource="http://localhost/tematres3.0/vocab/?tema=35"/>
</skos:Concept>
```
4.0 Concluding remarks

The concept is the fundamental element of the SKOS vocabulary. In basic SKOS, conceptual resources (concepts) can be identified with URIs, labelled with lexical strings in one or more natural languages, documented with various types of note, semantically related to each other in informal hierarchies and association networks and aggregated into concept schemes. Finally, semantic relations play a crucial role for defining concepts, by assigning meaning and context. (SKOS 2009).
Faceted classification was conceived by Ranganathan to deal with subject components of compound subjects; it has been evolving towards handling with different properties of documents – subjects, authors, publication date – and with different properties of things, organizing them in descriptive ontologies. Faceted classification is an information retrieval device. Its effectiveness depends on the analysis and identification of the properties of the objects being described, according to relevance criteria. While taxonomic classification emphasizes the Aristotelian substances, faceted classification emphasizes the accidents, i.e., the different and relevant properties (according to the modelling here proposed) through which the information objects may be retrieved.

The aim of this paper is to develop a conceptual model of a faceted classification and its components; this model is coded in SKOS, and the codification evaluated.

We propose conceptualizing facets as properties (relationships and attributes, attributes and characteristics, data type and object type properties, according to the different authors or sources identified). This modelling option makes explicit the components encompassed by a facet and its interrelations: a class with its properties, a facet (one of the properties of a class, a criterion) and the subclasses or instances created by applying that criterion to the class. Conceptualized in this way, its coding in SKOS follows. The coding maintains and adds to the constructs of the SKOS vocabulary, in addition to using similar constructions as those used in Semantic Web languages, such as RDF, RDFS and OWL, such as classes, subclasses, and properties etc. This helps to bring KOS closer to the mainstream of the Semantic Web.

Acknowledgments:
We are grateful to Prof. Linair Campos for the valuable comments about this paper. This work was carried out with the support of the Brazilian agencies CAPES - Financing Code 001 and CNPq, grant number 305253/2017-4.

References


© Ergon - ein Verlag in der Nomos Verlagsgesellschaft