Description Logic and Faceted Knowledge Representation

Uta Priss
School of Library and Information Science
Indiana University, Bloomington, IN 47405, USA
upriss@indiana.edu

1 Introduction

The term "facet" was introduced into the field of library classification systems by Ranganathan in the 1930's [Ranganathan, 1962]. A facet is a viewpoint or aspect. In contrast to traditional classification systems, faceted systems are modular in that a domain is analyzed in terms of baseline facets which are then synthesized. In this paper, the term "facet" is used in a broader meaning. Facets can describe different aspects on the same level of abstraction or the same aspect on different levels of abstraction. The notion of facets is related to database views, multicontexts and conceptual scaling in formal concept analysis [Ganter and Wille, 1999], polymorphism in object-oriented design, aspect-oriented programming, views and contexts in description logic and semantic networks.

This paper presents a definition of facets in terms of faceted knowledge representation that incorporates the traditional narrower notion of facets and potentially facilitates translation between different knowledge representation formalisms. A goal of this approach is a modular, machine-aided knowledge base design mechanism. A possible application is faceted thesaurus construction for information retrieval and data mining. Reasoning complexity depends on the size of the modules (facets). A more general analysis of complexity will be left for future research.

2 Faceted Knowledge Representation

The elements (or uniformities) of faceted knowledge representation are *units*, *relations* and *facets*. Units are atomic elements or tuples of atomic elements. Relations are sequences or matrices of 0's and 1's. They only obtain a meaning if they are applied to a domain (i.e., sets of units). Conceptual relations (roles) that are not binary or unary are modeled as higher level facets (see below) but are not "relations" in terms of the following definition.

Definition 1: $\mathcal{U}$ denotes a set of *uniformities*. Elements of U are denoted by lower case letters, subsets by upper case letters. $\mathcal{N}$ denotes a set of *units* with $\mathcal{N} \subseteq \mathcal{U}$. $\mathcal{R}$ denotes a set of *relations* with $\mathcal{R} \subseteq \mathcal{U}$. A unary relation is a sequence of 0's and 1's. A binary relation is a binary matrix, i.e., an array of 0's and 1's. All sets in this paper are finite.

A *facet* is a viewpoint or aspect of given uniformities and their relations. These can be constrained by rules, which are constructed from uniformities and operators. The following list of operators is only a suggestion and can be modified depending on the requirements of applications. For definitions of the relational operators compare, for example, [Pratt, 1992].

Set operators for units and sets of units: $\in$, $\subset$, $\subseteq$, $=$, Cartesian product ($\times$).
Relational operators: relational union ($\cup$), relational intersection ($\cap$), relational complement (^{c}); relational product ($\circ$) and its de Morgan complement ($\bullet$), relational inverse (^{d}); relational equality ($=$), relational containment ($\subset$ and $\subseteq$).

Relations are meaningless unless they are applied to sets of units as domain and codomain. This is formalized in basic facets.

Definition 2: A *basic facet* consists of a relation and a set of units as domain and, in the case of binary relations, a set of units as codomain. The notation is $f = (N; r)$ or $f = (N_1, N_2; r)$, respectively. For the units that correspond to a 1 in the sequence or matrix, the relation is written as $n \in r_f$ or nr_f and $(n_1, n_2) \in r_f$ or $n_1 r_f n_2$, respectively. The index "f" can be omitted in context. The set of basic facets (denoted by $\mathcal{F_B}$) is a subset of $\mathcal{U}$.

For the relational operators the usual equivalences follow: $n_1(r_1 \cup r_2) \Leftrightarrow n_1 r_1$ or $n_1 r_2$, and so on. It is always assumed that the sets of units are linearly ordered and, for example, in the case of the relational union, the sets are identical and identically ordered.

Basic facets are similar to formal contexts in formal concept analysis except that an interpretation as a concept lattice is only one possibility for a basic facet. For example, the relation can also be interpreted as an adjacency matrix of a graph. Basic facets are special kinds of facets, which are defined as follows.

Definition 3: A *facet* is a relational structure consisting of uniformities and/or sets of uniformities and rules that constrain the uniformities and that are formed using uniformities and operators. A facet f_1 that is used for constructing another facet f_2 is called *subfacet* of f_2, denoted by $f_1 \sqsubset f_2$. The set of facets (denoted by $\mathcal{F}$) is a subset of $\mathcal{U}$. The following conditions must be fulfilled:
a) $\sqsubset$ is acyclic and transitive.
b) Facets that do not contain subfacets are basic facets.
c) Only units, sets of units and relations of a facet or its subfacets can be used in rules.
d) For every relation in a facet there is at least one basic facet that contains that relation.

The *extension* of a facet refers to the units that are described by the facet. The *intension* of a facet refers to the mathematical structure formed by its relations. The precise method of calculating extension or intension depends on the application. Facets are called *extensionally equal* (*intensionally equal*) if they have equal extensions (intensions), respectively. Facets are equal if they are extensionally and intensionally equal. As an example, database queries can be modeled as facets. They are extensionally equal if they result in the same set of retrieved rows. They are intensionally equal if they are logically equivalent according to relational calculus. Renaming of units or relations may not change intensions, but usually changes extensions.

Interpretations are mappings from a set or powerset of uniformities to a set or powerset of uniformities. *Visualizations* are interpretations that map uniformities onto elements of a graphical representation. In faceted knowledge representation, interpretations are defined broader than in description logic. A mapping of uniformities onto elements of an external domain is a special kind of interpretation. Other interpretations allow conversion between different knowledge representation formalisms. Intension and extension of facets can be defined in terms of interpretations. *Meta-facets* are facets that describe construction methods or operators for sets of facets.

3 A Faceted Thesaurus as a Description Logical T-Box

Using faceted knowledge representation a description logical T-Box can be constructed from smaller T-Boxes in a modular manner. Figure 1 shows an example of a T-Box that is represented as a traditional faceted thesaurus. The generic relation (IS-A relation) forms an ordered set based on terms that can be aggregated (such as "professor" and "student") and composed (such as "full-time in-state student"). Each (sub-)facet is identified by its unique top term. If facets are constructed by term aggregation, the set of concepts equals the union of the sets of concepts of the subfacets. If facets are constructed by term composition, which is indicated by enclosing the subfacet top terms in angle brackets, the set of concepts is the direct product of the sets of concepts of the subfacets. Double angle brackets indicate roles, which are formally identical to facets created by term composition but are interpreted differently (see below). The @ symbol is a pointer and indicates that the facets are defined elsewhere. For more details on faceted thesauri compare [Priss and Jacob, 1999].

Each facet consists of two basic facets: $(N_t, N_t; r_g)$ which represents the generic relation among terms and $(N_t; r_b)$ which identifies term composition. The facet further consists of subfacets, rules for term composition and aggregation, and further rules, such as "Good-standing = International $\sqcap$ Full-time", which express constraints, such as "international students must be enrolled full-time to be in good standing".

```
person
      professor
      student
            <time>
                 full-time
                 part-time
            <residence>
                 in-state
                 out-of-state
                 international
course
      graduate
      undergraduate
teach
      <<professor @>>
      <<course @>>
```

Figure 1

In an interpretation over a domain, every term (concept) is mapped onto a unary relation and every role is mapped onto a binary matrix. An interpretation is a model for the thesaurus, if $n_1^{\mathcal{I}} \subseteq n_2^{\mathcal{I}} \Leftrightarrow n_1 r_g n_2$ for all terms $n_1, n_2 \in N_t$ and all rules are fulfilled. Existential and universal quantification correspond to the relational composition and its de Morgan complement, respectively. For example, queries (concept expressions) can be formulated, such as "X= Professor $\cap$ (TEACH $\circ$ Graduate)" for professors that teach at least one graduate level course and "Y= Professor $\cap$ (TEACH $\bullet$ Graduate)" for professors that teach only graduate level courses.

References

[Ganter and Wille, 1999] Ganter, Bernhard; Wille, Rudolf. *Formal Concept Analysis: Mathematical Foundations*. Springer Verlag, Berlin-Heidelberg, 1999.

[Pratt, 1992] Pratt, Vaughan. *Origins of the calculus of binary relations*. In: Proc. 7th Annual IEEE Symp. on Logic in Computer Science, Santa Cruz, CA, 1992, p 248-254.

[Priss and Jacob, 1999] Priss, Uta; Jacob, Elin. *Utilizing Faceted Structures for Information Systems Design*. Proc. of ASIS Annual Meeting, 1999.

[Ranganathan, 1962] Ranganathan, S. R. *Elements of library classification*. Asia Publishing House, Bombay, 1962, p 45-70.