HHS Public Access

Author manuscript

Proc IEEE Int Conf Inf Reuse Integr. Author manuscript; available in PMC 2017 April 14.

Published in final edited form as:

Proc IEEE Int Conf Inf Reuse Integr. 2015 August ; 2015: 388–395. doi:10.1109/IRI.2015.65.

Modeling Reusable and Interoperable Faceted Browsing Systems with Category Theory

Daniel R. Harris

Center for Clinical and Translational Science, University of Kentucky, Lexington, KY, USA

Abstract

Faceted browsing has become ubiquitous with modern digital libraries and online search engines, yet the process is still difficult to abstractly model in a manner that supports the development of interoperable and reusable interfaces. We propose category theory as a theoretical foundation for faceted browsing and demonstrate how the interactive process can be mathematically abstracted. Existing efforts in facet modeling are based upon set theory, formal concept analysis, and lightweight ontologies, but in many regards, they are implementations of faceted browsing rather than a specification of the basic, underlying structures and interactions. We will demonstrate that category theory allows us to specify faceted objects and study the relationships and interactions within a faceted browsing system. Implementations can then be constructed through a category-theoretic lens using these models, allowing abstract comparison and communication that naturally support interoperability and reuse.

Index Terms

Data models; interactive systems; software reusability; information architecture

I. Introduction

Faceted browsing (also called faceted search or faceted navigation) is an exploratory search model, where facets assist in the navigation of search results [1]. Facets are simply attributes attached to the actual objects being explored. An example of a facet attached to a book could be its genre or publication date. In a typical faceted browsing system, a user is shown search results alongside a list of related, relevant facets, allowing interactive filtering and expansion of results [2]. A faceted taxonomy is the collection of facets provided by the interface and is often organized as sets, hierarchies, or graphs.

The design of the system's underlying model directly impacts the user's ability to filter, rank, and interact with the facets; in fact, some models contain no interactivity [3]. Wei et al. observed three major theoretical foundations behind current research of facet models: set theory, formal concept analysis, and lightweight ontologies [3]. Facet modeling focuses on the formal representation of faceted data and the interactive consequences that follow when using that model.

The motivation for choosing category theory began when designing the next phase of DELVE [4], our framework for creating visualizations for browsing biomedical literature. Specifically, we encountered difficultly in modeling DELVE's ability to create numerous visualizations, which are either controlled by facets or contain faceted structures. Additionally, one visualization may impact another (either by filtering or focusing). For example, how can you effectively represent a hierarchical tree of facets which is simultaneously browsable and capable of spawning faceted graphs containing interactive, linked nodes? A model of faceted browsing that is capable of representing faceted taxonomies generically would enable the quick creation of interoperable faceted components within an interface and enable their reuse in either other parts of the interface or in a different interface altogether. Although this abstraction is possible with set theory, the notation quickly becomes cluttered and error-prone. It is very difficult to incorporate existing work on faceted browsing due to the vast variety of models and implementations. A modeling methodology that is capable of operating at a high level of abstraction is necessary.

Some faceted systems, such as hierarchical faceted categories [1], [5], are implemented without a true theoretical foundation [3]; in this context, categories refer to how facets categorically index items and is not related to category theory. In general, category theory aims to represent objects and relations at their most intrinsic, abstract level and is appropriate for modeling problems in the sciences [6], including computer science [7]. The volume of existing work for faceted browsing systems lends itself to a higher degree of abstraction, where existing works can become interoperable and reusable in new research settings. We will demonstrate that category theory is an appropriate framework for developing such abstractions by establishing facets and faceted taxonomies as categories.

II. Background

Faceted browsing systems enable effective use of faceted taxonomies and facet classification, the process of assigning facets to the resources to be queried. The utility of faceted classification and faceted taxonomies is well-understood [1], [2], [8], even as a pivotal element to modern information retrieval [9]. Faceted taxonomies can aid in the construction of information models [10] or aid in the construction of a larger ontology [11]. We focus on modeling faceted browsing in a way that enables the design of reusable and interoperable faceted taxonomies within the interface.

In order to fully understand our motivation, we must discuss existing faceted browsing models. When discussing these efforts, it is important to keep in mind that these models were constructed for a single system with a single faceted taxonomy; although each system was clearly an innovative and successful initiative, reusability and interoperability with future systems was not a priority or consideration discussed.

Additionally, basic knowledge of category theory is necessary to understand our model.

A. Foundations of Facet Models

Of the three major foundations of facet modeling, set theory efforts tend to provide a model that explicitly includes interactive operations, such as filtering and ranking [3]. Formal

concept analysis focuses on defining facets and faceted structures for knowledge [12] and has deep-seated roots in lattice theory in order to provide an organizational structure to faceted browsing. Light-weight ontologies provide an easy way to apply natural-language labels to concepts organized in an ontology [13], but do not explicitly model interactive operations.

Any implementation of faceted browsing, whether its foundation be grounded in set theory or lattice theory, could be abstracted into a category-theoretic framework as objects and relations. Because natural connections between category and set theory exist [6], [14], our work is most comparable to existing efforts in set theory. The facet modeling efforts that explicitly use set theory as a foundation differ in their core definitions and how they model filtering and ranking of facet objects: Dynamic Taxonomies [15] is a classic way of dynamically representing taxonomies with *is-a* relationships; Category Hierarchies are defined as connected, rooted directed acyclic graphs [16]; Generalized Formal Models [17] use entity-relationship diagrams to represent faceted hierarchies; FaSet [18] implements facets and queries as sets within relational databases. Each of these implementations have their own base definition of what it means to be a facet. In FaSet, a facet F is a set of items and if the system has multiple facets, they are disjoint: $F_a \cap F_b = \varnothing$ [18]. The model is then constructed axiomatically using the base definition of a facet.

Our work is also similar to efforts based on formal concept analysis [12] simply due to their shared abstract nature. We focus on modeling faceted structures once a representation for knowledge has been chosen, including lattices from formal concept analysis; we do not compete with formal concept analysis, but rather enable its reuse by providing a model capable of representing it consistently with other faceted structures. In other words, a lattice can peacefully coexist and interact with simpler structures such as sets and hierarchies. Many systems contain only one faceted taxonomy, but systems like DELVE can contain multiple visualizations; these visualizations either contain or are controlled by independent faceted taxonomies that may or may not share the same structure.

An example would be an interface that initially contains a visualization of a simple dynamic hierarchy depicting basic *is-a* relationships; for a given node, the interface could interactively allow one to visualize more complex relationships that are stored in a different knowledge structure, such as graphs or lattices.

Another example would be augmenting a faceted taxonomy with additional facets from an external faceted taxonomy. A more concrete example will help illustrate how such a situation can arrive naturally. At our local university hospital, our data warehouse provides a faceted interface which allows individuals to obtain aggregate counts of patients that match facets selected by the user; this allows quick feasibility checks of clinical research projects. Facets are arranged as a simple hierarchy and include items such as demographics (age, race, marital status) and vital signs (height, weight, bmi, blood pressure, heart rate, respiratory rate). These facets, illustrated in Fig. 1, are based on what the electronic medical records (EMRs) for our university hospital provides. In Fig. 1, lines represent relationships and ellipsis imply there are some relationships not shown for space considerations.

The EMR also uses drug codes that link to an external proprietary system where drugs are organized as a two-level hierarchy. For example, as seen in Fig. 2, buprenorphine *is-an* analgesic and an analgesic *is-a* central nervous system agent. We include this external hierarchy as part of our faceted taxonomy by adding drug as a facet. This augmentation enables one to search for classes of drugs, rather than just the drug codes found directly in the patient's EMR.

B. Category Theory

Category theory has been demonstrated to be practical and useful for modeling problems in the sciences [6], including physics [19], cognitive science [20], and computational biology [21]. In database theory, categories can model databases [6], [22] and can elegantly support data migration between schemas [23]. Additionally, ologs use category theory for representing knowledge and modeling real-world situations with the goal of enabling reusable, transferable, and comparable research [24]. Category theory can also be used in conjunction with semiotics as a foundation for information visualization [25], where the process of visualization forms a category.

Because our model leverages known categories, a working knowledge of category theory is needed. The definition of a category varies according to the style and notation adopted by its authors for their intended audience [6], [7]. Informally, a category $\mathscr{C}$ is defined by stating a few facts about the proposed category (specifying its objects, morphisms, identities, and compositions) and demonstrating that they obey identity and associativity laws.

Definition 1—A category $\mathscr{C}$ consists of the following:

1. A collection of objects, $Ob(\mathscr{C})$.
2. A collection of morphisms (also called arrows). For every pair $x, y \in Ob(\mathscr{C})$, there exists a set $Hom_{\mathscr{C}}(x, y)$ that contains morphisms from x to y [6]; a morphism $f \in Hom_{\mathscr{C}}(x, y)$ is of the form $f : x \rightarrow y$, where x is the domain and y is the codomain of f.
3. For every object $x \in Ob(\mathscr{C})$, the identity morphism, $id_x \in Hom_{\mathscr{C}}(x, x)$, exists.
4. For $x, y, z \in Ob(\mathscr{C})$, the composition function is defined as follows: $\circ : Hom_{\mathscr{C}}(y, z) \times Hom_{\mathscr{C}}(x, y) \rightarrow Hom_{\mathscr{C}}(x, z)$.

Given 1–4, the following laws hold:

1. identity: for every $x, y \in Ob(\mathscr{C})$ and every morphism $f : x \rightarrow y$, $f \circ id_x = f$ and $id_y \circ f = f$.
2. associativity: if $w, x, y, z \in Ob(\mathscr{C})$ and $f : w \rightarrow x$, $g : x \rightarrow y$, $h : y \rightarrow z$, then $(h \circ g) \circ f = h \circ (g \circ f) \in Hom_{\mathscr{C}}(w, z)$.

As an example, **Set** is the category whose objects are sets and whose morphisms are functions between sets [6], [7]. This implies that the set theoretic implementations of faceted browsing could also be directly abstracted and argued through category theory; we can comment on how our model could consume other models after we outline how our model works. The point of such structures is that they generalize sets by specifying families of

elements rather than single elements, a formalization which enables exploration of structural similarities [6].

An easy way to construct a new category is to modify an existing category and create a subcategory by taking a subset of its objects and morphisms. One can think of this as simply removing some of the objects from the original category, then removing any traces of the corresponding morphisms. The two categories behave similarly for those objects and morphisms found in both; in fact, injective monomorphisms exist between the sub-category and the larger category [7]. Identity and associativity laws are still obeyed by the remaining objects and their morphisms.

Definition 2—A category $\mathscr{B}$ is a subcategory [7] of $\mathscr{C}$ if

1. $Ob(\mathscr{B}) \subseteq Ob(\mathscr{C})$
2. for all $x, y \in Ob(\mathscr{B})$, $Hom_{\mathscr{B}}(x, y) \subseteq Hom_{\mathscr{C}}(x, y)$
3. the identity morphisms and compositions for $\mathscr{B}$ are copied from $\mathscr{C}$.

In our model, we will relate the concept of a facet and facet taxonomy to existing, well-known categories: **Rel** and **Cat**.

Definition 3—Rel is the category of sets as objects and relations as morphisms [7], where we define relation arrows $f : X \rightarrow Y \in Hom_{\mathbf{Rel}}(X, Y)$ to be a subset of $X \times Y$.

In other words, any subset of $X \times Y$ is a relation from X to Y. Any binary relation is allowed, but most examples demonstrate the utility of "<", "≤", and "⊆". This category uses the composition of relations instead of functional composition; if $f \in Hom_{\mathbf{Rel}}(X, Y)$ and $g \in Hom_{\mathbf{Rel}}(Y, Z)$, then $(x, y) \in g \circ f$ if and only if for some $y \in Y$, $(x, y) \in f$ and $(y, z) \in g$. The identity morphisms, $id_X \in Hom_{\mathbf{Rel}}(X, X)$, are the so-called diagonal relationships ($\{(x, x) | x \in X\}$). **Set** is actually subcategory of **Rel** [7].

Definition 4—Cat is the category of small categories. The objects of **Cat** are small categories and the morphisms are functors (mappings between categories).

There are several ways to define a small category, but the simplest is to say that a category $\mathscr{C}$ is small if $Ob(\mathscr{C})$ is a set where $\mathscr{C} \notin Ob(\mathscr{C})$ [6]. We informally defined functors as mappings between categories, but additional conditions are needed.

Definition 5—A functor F from category $\mathscr{C}_1$ to $\mathscr{C}_2$ is denoted $F: \mathscr{C}_1 \rightarrow \mathscr{C}_2$, where $F : Ob(\mathscr{C}_1) \rightarrow Ob(\mathscr{C}_2)$ and for every $x, y \in Ob(\mathscr{C}_1)$, $F : Hom_{\mathscr{C}_1}(x, y) \rightarrow Hom_{\mathscr{C}_2}(F(x), F(y))$. Additionally, the following must be preserved:

1. identity: for any object $x \in Ob(\mathscr{C}_1)$, $F(id_{\mathscr{C}_1}) = id_{F(\mathscr{C}_1)}$.
2. composition: for any $x, y, z \in Ob(\mathscr{C}_1)$ with $f : x \rightarrow y$ and $g : y \rightarrow z$, then $F(g \circ f) = F(g) \circ F(f)$.

Functors also play an important role in constructing the underlying graph of a category, which will be a key element of creating reusable facets and faceted structures.

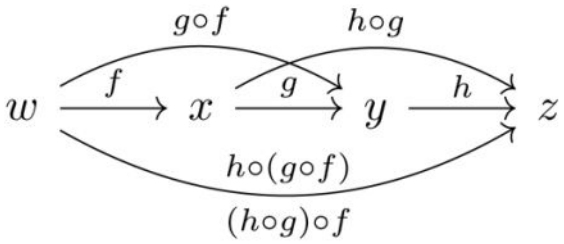
C. Why category theory?

We mentioned that category theory has been used to model several other practical problems, but we should comment on why it is also appropriate to model faceted browsing. The core issue is that faceted taxonomies come in many shapes and forms; this heterogeneity, while a sign that new and novel systems are being developed, is counterproductive for reusing faceted information effectively. Category theory provides the language for reasoning with these diverse structures in a consistent and productive environment. Once a category is defined, it becomes a bed for computations, such as transformations and products. As an example, consider n-ary products of objects within a category [7], which act as a categorical version of n-ary Cartesian products.

Definition 6—The n-ary product of a list $A_1, A_2, \ldots, A_n$ containing n objects (not necessarily distinct) of a category is an object A with morphisms $p_i : A \rightarrow A_i$ for $i = 1, \ldots, n$. This is denoted as $A_1 \times A_2 \times \ldots \times A_n$ or simply $\prod_{i=1}^{n} \mathbf{A}_i$ [7].

We will demonstrate later that n-ary products model faceted universes and faceted queries. Beyond products, category theory is conceptually consistent with faceted browsing in that relations within a faceted taxonomy naturally mimic the constraints of categories. Consider a faceted taxonomy with *is-a*-relations and observe that the following analogies hold:

1. An object has an identity function: x *is-a* x.
2. Relations can be composed: (x *is-a* y) *is-a* z.
3. Commutative diagrams typically demonstrate how objects and morphisms in a category obey associativity:



The same works for *is-a*-relations: ((*x is-a y*) *is-a z*) is equivalent to (*x is-a* (*y is-a z*)).

Given this discussion, there are clear benefits in choosing category theory as a modeling foundation, but we must acknowledge that the learning curve can be an impediment of adoption by those less familiar with the theory. We argue that any unfamiliar theory can be difficult to learn and that the benefits outweigh the obstacles.

III. A Category-theoretic Model

A faceted system is comprised of many implicitly intertwined parts: facets, a taxonomy that organizes the facets, an ability to select or focus on certain facets, and an ability to present

the results of a selection or faceted query in an effective manner [3]. Each of these components and their extensions can be abstractly modeled with category theory.

A. Facet Types

We wish to be as general as possible in our abstractions so that any system with any faceted taxonomy can be modeled, regardless of the particular nuances of the facets and intra-facet relationships. The faceted taxonomy presented in an interface can contain several unrelated (or disjoint) sub-facets. For example, a book's price typically has nothing to do with its genre, and the construction and maintenance of the corresponding taxonomic structures are completely independent. We can refer to a facet, such as price and genre, as a *facet type*. The interface typically presents facets underneath a heading that indicates its type.[1] Within a facet type, facets are directly relatable and comparable. In other words, for our example, prices relate to prices and genres relate to genres; "$10–$20" can be a child of " <$100", but has no relationship with "horror". Fig. 1 shows a facet type for patient demographics; Fig. 2 shows a facet type for medications. Demographics and medications are clearly disjoint: no taxonomic relations exist between these two types.

Definition 7—A facet type (a facet i and its related sub-facets) of a faceted taxonomy is a sub-category of **Rel**, the category of sets as objects and relations as morphisms. Let's call this sub-category $\mathbf{Facet}_i$ and let $Ob(\mathbf{Facet}_i) = Ob(\mathbf{Rel})$ and let the morphisms be the relations that correspond only to the $\subseteq$-relations. The identity and composition definitions are simply copied from **Rel**.

In other words, $\mathbf{Facet}_i$ is just a slimmer version of **Rel**, where we know exactly what binary relation is being used to order the facets. If a sub-category contains all objects of its parent, it is said to be wide [7]. In fact, $\mathbf{Facet}_i$ is said to be a representative subcategory of **Rel** because it contains an isomorphic copy of every object of **Rel** [7]. The other relations that could possibly be represented by **Rel**, such as "<" and "≤", are meaningless for categorically organizing concepts into a taxonomy. The objects of $\mathbf{Facet}_i$ are sets; they represent abstract collections of resources that have been classified to belong to that facet through faceted classification. We do not need to distinguish between individual resources within these sets because they all categorically behave the same: they either belong to facet or they do not. Between the objects of $\mathbf{Facet}_i$, morphisms exist that dictate their taxonomic relationship.

As a concrete example, if $\mathbf{Facet}_{Med}$ is the medication facet-type in Fig. 2, then {"central nervous system agent", "analgesic", "buprenorphine", …} $\in Ob(\mathbf{Facet}_{Med})$. The morphisms of $\mathbf{Facet}_{Med}$ dictate the relationships between objects. Suppose that object x is the set for "analgesic" and object y is the set for "central nervous system agent", then there exists a morphism $f : x \rightarrow y \in Hom_{\mathbf{Facet}_{Med}}(x, y)$, meaning "analgesic" *is-a* subset of "central nervous system agent".

The $\subseteq$-relation is powerful for specification: it allows for facets to be ordered by inclusion, which can model any structure where x is categorically related to y; this is a pivotal

[1]The interface may wish to include the name of the type as a selectable facet too. This meta-facet is mostly an organizational tool that aids in drawing the faceted taxonomy.

component to most faceted implementations. We intentionally use the words specification and implementation; our category-theoretic model specifies the fundamental objects and relationships that our implementation can utilize. For example, we can implement a tree that only has *is-a*-relationships since it is possible to order the facets by inclusion. We will show how this type of tree, which is heavily related to dynamic taxonomies [15], can be represented with category theory in section IV. In a more complicated example, one could construct a visualization with a graph, where the current facet is recursively drawn connected to a subset of its subsumptive facets.

B. Focused Selections

Given a facet, we need to describe how any selection within the facet can be modeled.

Definition 8—A subcategory $\mathbf{Focus}_i \subseteq \mathbf{Facet}_i$[2] can be constructed to represent a focused selection for $\mathbf{Facet}_i$.

We simply discard any undesirable objects (and their corresponding morphisms) to create a new category that represents a focused collection of facets. The identity and composition functions can be copied from $\mathbf{Facet}_i$. Selecting objects in facet type is the simplest form of interacting with a faceted browser. As a concrete example, if $\mathbf{Facet}_{Med}$ is the medication facet-type in Fig. 2, then one possibility is {“central nervous system agent”, “analgesic”} $\in Ob(\mathbf{Focus}_{Med})$. There is no limit on the number of objects kept or discarded; it is possible to discard all objects or to keep all objects, although practical limitations might stem from the user’s ability to interact with the interface.

C. Faceted Taxonomies

Presentation of facets varies according to the interface’s design, but facets are commonly presented as flat lists (through widgets such as radio buttons, check-boxes, etc) and hierarchies. Hierarchies, if restricted to be non-overlapping, are often represented as a tree where each facet is limited to having one parent. A non-hierarchical (or flat) facet is a special case of a hierarchical facet: it is simply a hierarchical taxonomy with one level. For example, if authors of books were a primitive facet, it could be represented as a hierarchy with author as the root and individual names as children. More complicated faceted interfaces may present taxonomies via visualizations, including graphs and network diagrams.

Definition 9—Let **FacetTax** be a category that represents a faceted taxonomy, whose objects are the disjoint union of $\mathbf{Facet}_i$ categories. In other words, let $\coprod_{i=1}^{n} \mathbf{Facet}_i$, where $n = |Ob(\mathbf{FacetTax})|$. The morphisms of **FacetTax** are functors (mappings between categories) of the form $Hom_{\mathbf{FacetTax}}(\mathscr{C}, \mathscr{D}) = \{F : \mathscr{C} \rightarrow \mathscr{D}\}$.

This disjoint union is precisely how we can merge facets from Fig. 1 and Fig. 2 into a single faceted taxonomy, *patients*, as illustrated in Fig. 3. To save space, not all objects and

[2] $A \subseteq B$ is commonly used denote that A is a subcategory of B, despite neither A nor B being an actual set [6].

relations are drawn. We will not formally prove that **FacetTax** is a true category because it is simply a sub-category of **Cat**, the category of categories.

D. Universe and Queries

The complexity of a faceted system naturally varies by the interface's design, but typically includes the ability to select (or focus) and de-select (or negate) facets within a facet type. The collective effort across all facets is then used to filter the faceted knowledge being presented.

Definition 10—A facet universe, U, is the n-ary product [7] within the **FacetTax** category, defined as $\prod_{i=1}^{n} \mathbf{Facet}_i$, where $n = |Ob(\mathbf{FacetTax})|$. The n coordinates of U are projection functors $P_j : \prod \mathbf{Facet}_i \rightarrow \mathbf{Facet}_j$, where $j = 1, \ldots, n$ is the jth projection of the n-ary product.

Note that since $\mathbf{Focus}_i \subseteq \mathbf{Facet}_i$, there exists a restricted universe $U_{\subseteq} \subseteq U$ where every facet is potentially reduced to a focused subset. The act of querying the universe is essentially constructing this restricted universe $U_{\subseteq}$.

Definition 11—A faceted query, Q, is the modified n-ary product [7] within the **FacetTax** category, defined as $\prod_{i=1}^{n} \mathbf{Focus}_i$, where $n = |Ob(\mathbf{FacetTax})|$. The n coordinates of Q are similarly defined as projection functors $P_j : \prod \mathbf{Focus}_i \rightarrow \mathbf{Focus}_j$.

A high-level overview of the interactions between $\mathbf{Facet}_i$, $\mathbf{Focus}_i$, and queries with the **FacetTax** category is illustrated in Fig. 4. To summarize, a faceted taxonomy is a category which contains **Facet** categories as objects; each **Facet** category contains objects and taxonomic relationships between the objects. Intuitively, focused selections are contained within their larger, unfocused facet categories. The universe of possible facets provided by the interface is the product across all **Facet** objects; a faceted query is simply a focused product chosen by the user from the universe of facets.

IV. Reusable Faceted Taxonomies

In our model, categories are acting as generic faceted structures, but in practical interface development, more is sometimes needed to support the range of possible designs and interactions. The generic nature of the morphisms of **FacetTax** allow it to abstractly represent any faceted taxonomy structure since the morphisms are simply $\subseteq$-relations.

We demonstrated that n-ary products were one useful computation enabled by our model, but we can also demonstrate how our base structure can transform to support different faceted structures, such as lists, hierarchies, trees, graphs, and lattices. This transformation can be an active and engaging element of the interface: a selected element from a basic list could render a graph of deeper relationships. Using Fig. 3 as an example, imagine that a preview of high level facet types are given in the form of a list (*demographics, medications,* …) and interactively selecting one of these higher level items results in the rendering of its remaining taxonomy. For example, clicking on *demographics* could draw the descendants of that object from the taxonomy in the interface.

Intelligent previews are a possible solution to the problem of having too little screen space to fully display an interface's facets, which is an open issue in faceted browsing research [3]. We can reuse the same facets, while manipulating their relations to fit other structures. These structures can be put into the same abstract framework to support interoperability between systems or between parts within a system. When designing a system where different structures interact with one another, the notation and representations become cumbersome. The burden of writing consistent abstractions with different structures is removed by using category theory.

A. Underlying Graphs

In order to show how graphs relate and interact with our model, we must formally define the category of graphs.

Definition 12—Grph is the category with graphs as objects [6]. A graph G is a sequence where $G := (V, A, src, tgt)$ [6] with the following:

1. a set V of vertices of G
2. a set A of edges of G
3. a source function $src : A \rightarrow V$ that maps arrows to their source vertex
4. a target function $tgt : A \rightarrow V$ that maps arrows to their target vertex

The graph underlying a category $\mathscr{C}$ is defined as a sequence $U(\mathscr{C}) = (Ob(\mathscr{C}), Hom_{\mathscr{C}}, dom, cod)$ [6] where $dom, cod : Hom_{\mathscr{C}} \rightarrow Ob(\mathscr{C})$ are domain and codomain functions. Note that $U(\mathscr{C}) \in Ob(\mathbf{Grph})$ and there exists a functor between categories $\mathcal{F} : \mathscr{C} \rightarrow \mathscr{D}$ that can create a graph morphism $U(\mathcal{F}) : U(\mathscr{C}) \rightarrow U(\mathscr{D})$. This works at two levels for our model: for **FacetTax** and for each **Facet** category. Given that there exists a functor U: **Cat** → **Grph**, **FacetTax** can produce graphs of $\mathbf{Facet}_i$ categories for $i = (1, \ldots, |Ob(\mathbf{FacetTax})|)$. Because every category has an underlying graph, each individual **Facet** can also be represented as a graph, where the objects are vertices and the morphisms are arrows. Although the taxonomy in Fig. 3 is likely best visualized as a simple hierarchy, other taxonomies might be more naturally suited to be represented as a graph. One possible use case is where graph algorithms might play a key role in the interface's design.

The key benefit to modeling our faceted taxonomy this way is that we can reuse the facets and reframe their morphisms to fit our needs; if we need to arrange the facets as graphs, we can do so. The sets of resources that the objects of **Facet** represent remain unchanged; resources are still classified with their facets. This embedded, faceted information can be reused through an alternate lens made possible by our model. In this case, the lens is a graph that was algebraically constructed, but other structures are also supported.

B. Sets

Rel is closely related to **Set**; both categories have sets as objects. In fact, it can be demonstrated that **Set** is a subcategory of **Rel** [7] and we will utilize this notion to show that **FacetTax** is compatible with **Set**. We can construct a functor $C : F \rightarrow S$ that maps **Facet** to **Set** by taking every set to itself and every relation to be a function from F to T. Similar to

how we can construct graphs, we are also free to leverage **Set** categories and can construct basic sets. This is helpful when a basic list needs to be constructed from our taxonomy, such as in the case of the previews discussed in the first part of Sec. IV.

C. Other Structures

Additionally, **Rel** is capable of representing lattices when ordering by inclusions [7] and a similar result can be obtained with our **Facet** categories. Again, a functor is how we can compute such a mapping between categorical representations. Category theory is providing two types of computation: within a category with concepts, such as products, and across categories with concepts, such as functors.

D. Existing Facet Models

We discussed several existing modeling efforts in our background section and we can discuss how such models would change and manifest under our category-theoretic model.

Dynamic taxonomies [15] use set theory as a theoretical foundation to help model facets. The model constructs taxonomies with *is-a* relationships by dynamically calculating the deep extension of a node:

$$\text{deepExtension}(C)=\{d|d \in \text{shallowExtension}(C^{'})\wedge(C^{'}=C\vee C^{'} \text{ is a descendant of C})\}$$

The shallow extension of C contains the direct descendants of C. Both shallow extension and *descendant-of* relationships are expressible as binary relations and, in particular, are expressible with the binary relations of **FacetTax**'s **Facet** categories. The shallow extension of an object $y \in Ob(\mathbf{Facet}_i)$ is the domain of the relations of $Hom_{\mathbf{Facet}_i}(x, y)$, e.g. all x Facet_i that are subsets of y. Nesting gives us the deep extension: the deep extension of an object $y \in Ob(\mathbf{Facet}_i)$ is the domain of the relations from $Hom_{\mathbf{Facet}_i}(z, Hom_{\mathbf{Facet}_i}(x, y))$ for any $x, y, z \in Ob(\mathbf{Facet}_i)$.

Category hierarchies are another example of facet models with set theory as a foundation [16]. In this model, category hierarchies are defined as connected and rooted directed acyclic graphs; the word category is unrelated to category theory in this context. Facets are defined as subgraphs of the category hierarchy. These subgraphs are exactly like the underlying graphs of categories discussed earlier. Once our facets are in a graph, the other features of their model follow.

Our goal in discussing how our model relates to existing work is to demonstrate how dynamic our category theory foundation can be and that it is able to drive interoperability and reuse. We can reuse existing models and merge them together by putting them in a common language. In this case, theory has direct consequences for specification and can help drive implementation.

E. Implementation

Even elegant abstractions can be rendered useless if they are not easily implementable. Category theory is strongly related to functional programming; one can even show that a

functional programming language forms a category of types and operations [7]. Furthermore, the use of objects (and morphisms between objects) as our model's foundation means implementation could be in any programming language that supports an object-oriented paradigm. For example, Scala [26] is a multi-paradigm language which supports both functional and object-oriented paradigms. In fact, the category of finite sets, often called **Fin** [6], [7], is easily implementable in Scala [27]; there are Scala libraries available that provide infrastructure for building categories [28], [29].

We contend that modeling has direct consequences for implementation: an inconsistent model yields an inconsistent system. If we cannot abstractly represent faceted browsing in an effective manner, there is little hope to extend and improve such a system and furthermore, the system cannot be adopted or easily modified by others. Theory must inform practice; we can use our abstractions to build stronger interfaces that support interoperability and reuse.

F. Computational Complexity

We have focused on structural complexity between categories and morphisms, but computational complexity should also be addressed. This only makes sense in the context of using abstract categories to write algorithms, which is one way our models can be applied and reapplied. An example is computing the transitive closure of **Fin**, which can be solved as the repeated accumulation of pairs of paths and can be demonstrated as having a complexity of $O(N^3 \log N)$ where N is the number of nodes [30]. Computational category theory [30] connects functional programming with category theory to bridge the gap between theory and implementation. A byproduct of computational category theory is that computational complexity can be studied.

V. Future Work

We are applying our model to the next stage of our interface and framework, DELVE [4], in order to represent faceted structures that help create or control other faceted structures. Our model, together with category theory, can help inform how to build a proper application programming interface (API) for faceted browsing. In this vein, one can mathematically prove that something is possible before implementation. The role of computational category theory in developing this API is an interesting unknown.

We have demonstrated that it is possible to represent facets, faceted taxonomies, and faceted queries with category theory. Once the faceted query is performed, the interface must let the user successfully interact and engage with the faceted information being presented. We wish to continue our model to include this exploratory search phase of faceted browsing.

For example, if the interface was designed to fulfill Shneiderman's information seeking mantra [31], how can tasks such as overview, filter, zoom, and details-on-demand be modeled? How can additional tasks from Shneiderman's task taxonomy [31], such as seeing relations, history, and extraction, be modeled?

Furthermore, if each of these tasks or interactions can be abstracted, does this model suggest anything about being able to measure the usability of the system? Usability is most commonly measured qualitatively, but we would like to explore the possibility of quantitatively measuring usability.

VI. Conclusions

We have established facets and faceted taxonomies as categories and have demonstrated how the computational elements of category theory, such as products and functors, extend the usefulness of our model.

The utility of faceted browsing systems is well-established in the digital libraries research community [8], [32], but current efforts could benefit from a more abstract framework that encourages reuse and interoperability. In this context, reuse and interoperability are at two levels: between systems and within a system. Our model works at both levels by leveraging category theory as a common language for representation and computation. Without this common language, it is difficult to abstractly model a system that utilizes multiple faceted structures (hierarchies, trees, graphs, lattices), even if there are shared notations and definitions between them.

We have demonstrated that category theory can be used to model faceted browsing and that it offers a consistent view of facets as objects and morphisms between objects. With our general framework for communicating mathematically about facets at a high level of abstraction, we can begin to construct interoperable interfaces and reuse existing efforts intelligently.

Acknowledgments

The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR000117. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work would not be possible without support from Drs. Jerzy W. Jaromczyk, Todd R. Johnson, and Ramakanth Kavuluru.

References

1. Hearst MA. Clustering versus faceted categories for information exploration. Communications of the ACM. 2006; 49(4):59–61.
2. Hearst, MA. SIGIR workshop on faceted search. ACM; 2006. Design recommendations for hierarchical faceted search interfaces; p. 1-5.
3. Wei B, Liu J, Zheng Q, Zhang W, Fu X, Feng B. A survey of faceted search. Journal of Web engineering. 2013; 12(1–2):41–64.
4. Harris, DR., Kavuluru, R., Yu, S., Theakston, R., Jaromczyk, JW., Johnson, TR. Proceedings of the Summit on Clinical Research Informatics. AMIA; 2014. Delve: A document exploration and visualization engine; p. 179
5. Yee, KP., Swearingen, K., Li, K., Hearst, M. Proceedings of the SIGCHI conference on human factors in computing systems. ACM; 2003. Faceted metadata for image search and browsing; p. 401-408.
6. Spivak, DI. Category Theory for the Sciences. MIT Press; 2014.
7. Barr, M., Wells, C. Category theory for computing science. Prentice Hall; New York: 1990.

8. Fagan JC. Usability studies of faceted browsing: a literature review. Information Technology and Libraries. 2013; 29(2):58–66.
9. Dawson, A., Brown, D., Broughton, V. Aslib proceedings. Vol. 58. Emerald Group Publishing Limited; 2006. The need for a faceted classification as the basis of all methods of information retrieval; p. 49-72.
10. Chu, H-J., Chow, R-C. An information model for managing domain knowledge via faceted taxonomies. Information Reuse and Integration (IRI), 2010 IEEE International Conference on; IEEE; 2010. p. 378-379.
11. Prieto-Díaz, R. A faceted approach to building ontologies. Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference on; IEEE; 2003. p. 458-465.
12. Priss, U. Faceted information representation. Proceedings of the 8th International Conference on Conceptual Structures; Springer-Verlag; 2000. p. 84-94.
13. Giunchiglia, F., Marchese, M., Zaihrayeu, I. Journal on data semantics. Springer; 2007. Encoding classifications into lightweight ontologies; p. 57-81.
14. Blass, A. Mathematical applications of category theory. Vol. 89. American Mathematical Society; 1984. The interaction between category theory and set theory; p. 84-84.
15. Sacco, GM., Tzitzikas, Y. Dynamic taxonomies and faceted search: theory, practice, and experience. Springer; 2009.
16. Li, C., Yan, N., Roy, SB., Lisham, L., Das, G. Proceedings of the 19th international conference on World wide web. ACM; 2010. Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia; p. 651-660.
17. Clarkson, EC., Navathe, SB., Foley, JD. Proceedings of the 9th ACM/IEEE Joint Conference on Digital libraries. ACM; 2009. Generalized formal models for faceted user interfaces; p. 125-134.
18. Bonino, D., Corno, F., Farinetti, L. Faset: A set theory model for faceted search. Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology; IEEE; 2009. p. 474-481.
19. Coecke, B., Paquette, ÉO. New Structures for Physics. Springer; 2011. Categories for the practising physicist; p. 173-286.
20. Phillips S, Wilson WH. Categorial compositionality: A category theory explanation for the systematicity of human cognition. PLoS computational biology. 2010; 6(7):e1000858. [PubMed: 20661306]
21. Spivak DI, Giesa T, Wood E, Buehler MJ. Category theoretic analysis of hierarchical protein materials and social networks. PLoS One. 2011; 6(9):e23911. [PubMed: 21931622]
22. Spivak DI. Simplicial databases. 2009 arXiv preprint arXiv:0904.2012.
23. Spivak DI. Functorial data migration. Information and Computation. 2012; 217:31–51.
24. Spivak DI, Kent RE. Ologs: a categorical framework for knowledge representation. PLoS One. 2012; 7(1):e24274. [PubMed: 22303434]
25. Vickers P, Faith J, Rossiter N. Understanding visualization: A formal approach using category theory and semiotics. Visualization and Computer Graphics, IEEE Transactions on. 2013; 19(6): 1048–1061.
26. Odersky, M. Technical Report. Programming Methods Laboratory, EPFL; Lausanne, Switzerland: 2007. The scala language specification, version 2.4.
27. Seeberger, H. Introduction to category theory in scala. Nov. 2010 retrieved January 25, 2015 from http://hseeberger.github.io/blog/2010/11/25/introduction-to-category-theory-in-scala/
28. Hupel, L. scalaz: functional programming for scala. 2014. retrieved January 25, 2015 from http://typelevel.org/projects/scalaz/
29. Hupel, L. scalaz.category api documentation. 2014. retrieved January 25, 2015 from http://docs.typelevel.org/api/scalaz/stable/7.1.0-M3/doc/#scalaz.Category
30. Rydeheard, DE., Burstall, RM. Computational category theory. Prentice Hall; 1988.
31. Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. Visual Languages, 1996. Proceedings., IEEE Symposium on; IEEE; 1996. p. 336-343.

32. Niu, X., Hemminger, B. Analyzing the interaction patterns in a faceted search interface. Journal of the Association for Information Science and Technology. 2014. [Online]. Available: http://dx.doi.org/10.1002/asi.23227

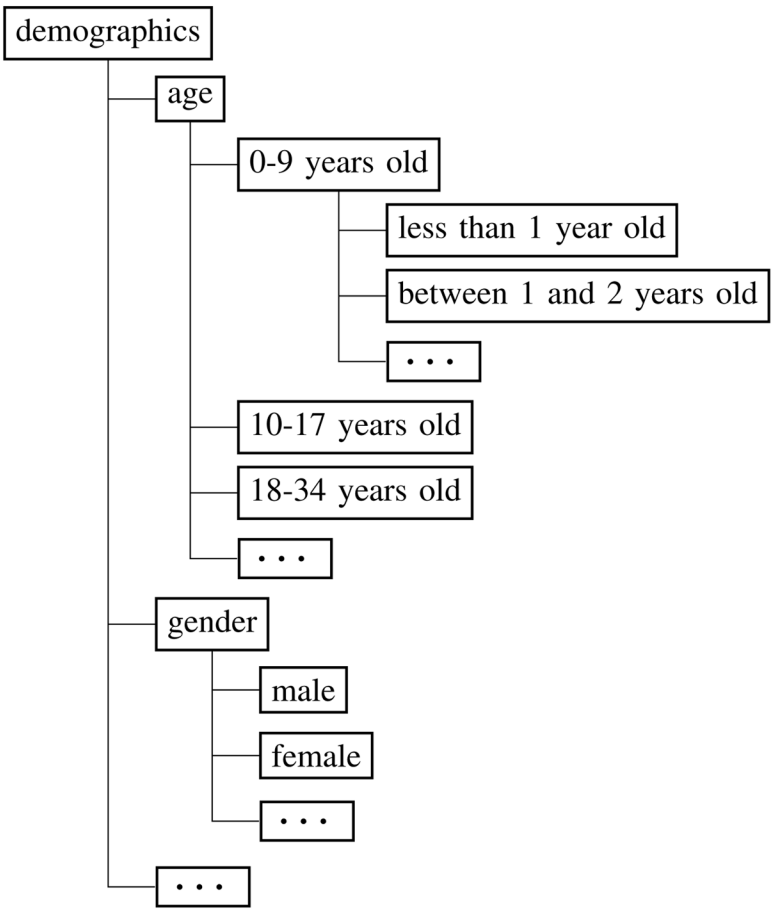


Fig. 1.

A small sample of facets available for patient medical records. For example, selecting female would query for all patients that are female.

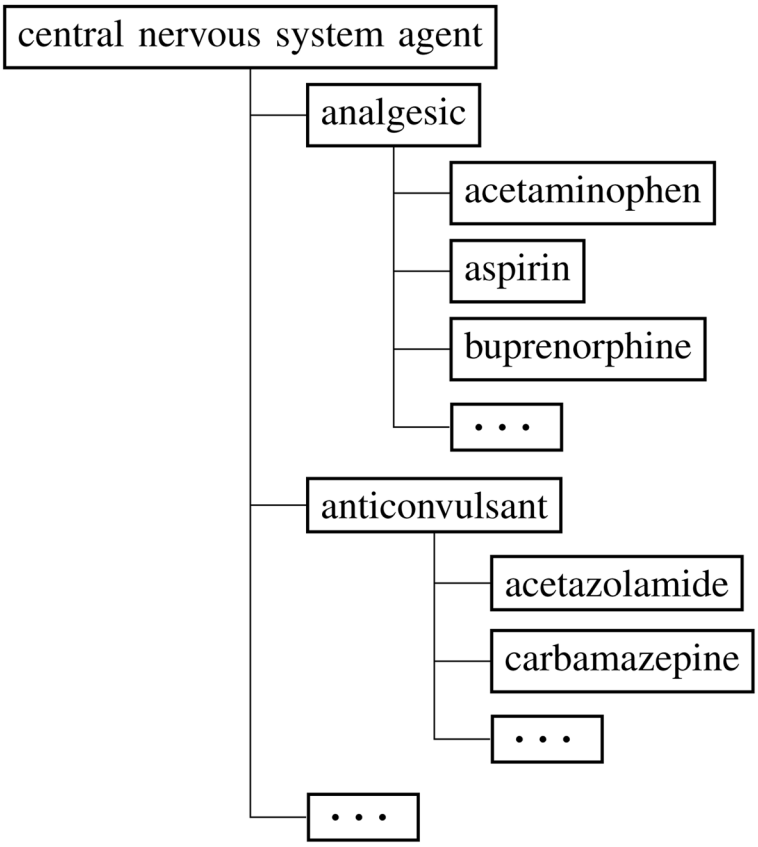


Fig. 2.
A small sample of facets available for drug administration records. For example, selecting analgesic would query for all records containing analgesics.

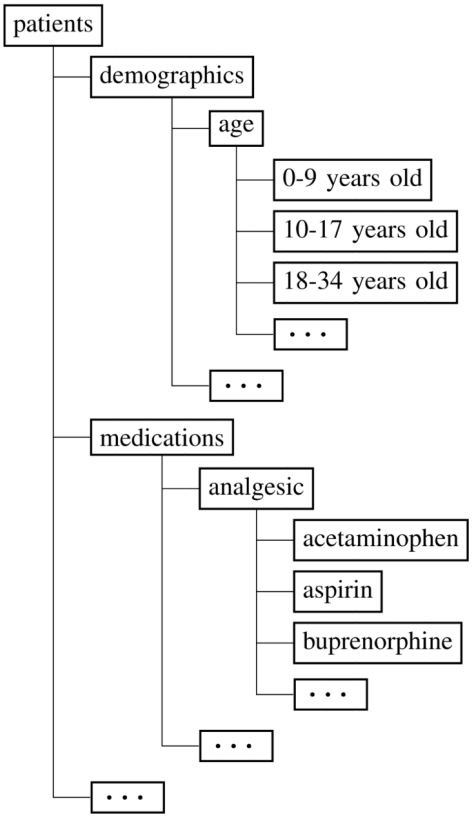


Fig. 3.

A faceted taxonomy, called patients, containing facet types of demographics and medications.

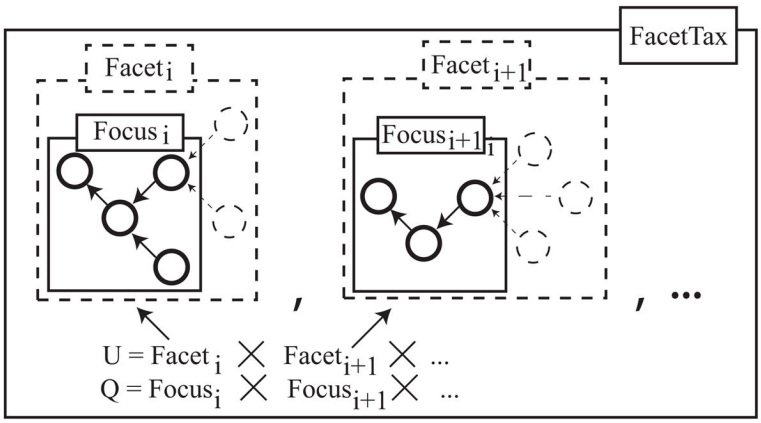


Fig. 4.

A high-level overview of **FacetTax**, **Facet**$_i$, **Focus**$_i$, and queries with focused subcategories.