1. Introduction

In this article we present a rigorous axiomatization of the notion of ontology from computer and information sciences, using terminology from the mathematical discipline of category theory. Specifically, we cast an ontological schema as a geometry in the sense of Lurie (see [?]), and the instances of a particular ontology as a structured space. Geometries are closely related to Grothendieck sites, which are of fundamental importance in category theory and algebraic geometry.

One purpose of this paper is to provide another link between information science and category theory, by connecting together important notions of the two fields. Both geometries in mathematics and ontologies in information science are structures which are designed to provide a “world of study” by specifying the requisite vocabulary and axiomatic rules that describe and govern that world. We make the similarities between these two notions clear by formulating ontological ideas in category-theoretic language.

In information science literature, the notion of ontology is kept deliberately vague. Different types of problems require different types of ontologies, and it may be best not to pin down a precise definition. In this paper we hope to show that geometries are at least strongly related to any reasonable notion of ontologies. In so doing, we aim to give a jumping-off point for those who wish to either use mathematical theorems and language in their study of ontologies or for those who wish to explicate and clarify distinctions between sites and ontologies.

Another purpose of this paper is to define transformations of ontologies, by directly importing the definition from Lurie’s work. In so doing, we provide a new way to deal with the problem of ontology integration. Given two ontologies, there may or may not be a transformation between them, but if they ostensibly “should” be related, then one will probably be able to find a morphism between sub-ontologies (ontologies either with fewer types or fewer rules [?]). Having a precise definition of
morphism is an important step toward solving the problems of integrating ontologies.

Finally, we relate databases to ontologies by producing a functor which sends the category of simplicial databases \([\mathcal{C}]\) to the category of ontologies. We characterize the image of this functor within the category of geometries. (An ontology is of relational-database type if there exists a set of “basic attributes” \(t_1, \ldots, t_n\), each of which having a distinguished cover \(v_{i,1}, \ldots, v_{i,m_i} \rightarrow t_i\) by admissible morphisms such that fiber products \(v_{i,j} \times t_i \rightarrow v_{i,k}\) are empty (i.e. the empty set forms a cover) for \(j \neq k\), and such that every object in \(t\) admits a unique admissible map to a product of basic attributes.)

The structure of the paper is as follows. In Section 2, we discuss the basics of ontologies and give some examples. In Section 3 we define Grothendieck sites, sheaves, geometries, and structured spaces. In Section 4 we explain how all of these notions connect by giving a dictionary relating their respective vocabularies, and by providing examples. In Section 5 we describe the relationship between ontologies and databases. Finally, in Section 6 we discuss transformations of geometries and show their applicability to ontology integration problems.

2. Ontologies

3. Geometries

We begin by recalling the definition of a site.

**Definition 3.0.1.** A site is a pair \((\mathcal{C}, J)\), where \(\mathcal{C}\) is a category and \(J\) is a function that assigns to each \(C \in \text{Ob}(\mathcal{C})\), a set \(J(C)\) in which each element \(T \in J(C)\) is itself a set \(T = \{t_{\alpha} : c_{\alpha} \rightarrow C\}\) of morphisms in \(\mathcal{C}\) with target \(C\); these data are required to satisfy axioms which we present after defining some terms. We call the sets \(T \in J(C)\) trait classifications on \(\mathcal{C}\) and individual morphisms \(t \in T\) trait classes on \(\mathcal{C}\).

(1) For each trait class \(t : c \rightarrow C\) on \(\mathcal{C}\) and each map \(D \rightarrow C\) in \(\mathcal{C}\), the fiber product \(c \times_C D\) exists in \(\mathcal{C}\).

(2) For any map \(f : D \rightarrow C\) in \(\mathcal{C}\) and any trait classification \(T \in J(C)\), the pullback \(f^*T = \{g^*(t)\mid t \in T\}\) is a trait classification on \(D\).

(3) If \(T = \{t_{\alpha} : c_{\alpha} \rightarrow C\}\) is a trait classification on \(\mathcal{C}\) and for each \(t_{\alpha} \in T\) the set \(U_{\alpha} = \{u_{\alpha,\beta} : b_{\alpha,\beta} \rightarrow c_{\alpha}\}\) is a trait classification on \(c_{\alpha}\), then the family of composites

\[
\bigcup_{t_{\alpha} \in T} \left\{ b_{\alpha,\beta} \right\}
\]

is a trait classification on \(\mathcal{C}\).

(4) If \(f : B \rightarrow C\) is an isomorphism in \(\mathcal{C}\) then the singleton set \(\{f\}\) is a trait classification on \(\mathcal{C}\).

**Definition 3.0.2.** Let \((\mathcal{C}, J)\) denote a site, \(X\) a topological space, and \(\text{Shv}(X)\) the category of sheaves of sets on \(X\). A functor \(F : \mathcal{C} \rightarrow \text{Shv}(X)\) is called topological if it preserves the finite limits which exist in \(\mathcal{C}\) and if, for every trait classification \(T \in J\), the induced map of sheaves

\[
\prod_{(t_{\alpha} : c_{\alpha} \rightarrow C) \in T} F(c_{\alpha}) \longrightarrow F(C)
\]
is surjective.

Given a pair of topological functors \( F, G : \mathcal{C} \to \text{Shv}(X) \), we will say that a natural transformation \( a : F \to G \) is topological if, for every trait class \( c \to \mathcal{C} \), the induced diagram of sheaves

\[
\begin{array}{ccc}
F(c) & \longrightarrow & G(c) \\
\downarrow & & \downarrow \\
F(C) & \longrightarrow & F(C)
\end{array}
\]

is a pullback square in \( \text{Shv}(X) \).

**Example 3.0.3.** If \( X \) is a point, then \( \text{Shv}(X) = \text{Sets} \) is the category of sets. A functor \( F : \mathcal{C} \to \text{Sets} \) is topological if it preserves finite limits and takes trait classifications to surjections.

I think that if \( X \) is sober, then a functor \( F : \mathcal{C} \to \text{Shv}(X) \) is local if, for each point \( x \in X \), the induced functor \( F_x : \mathcal{C} \to \text{Sets} \) on stalks (given by composition of \( F \) with \( \text{colim}_{x \in U} - (U) : \text{Shv}(X) \to \text{Sets} \)) is topological.

**Definition 3.0.4.** Let \( F : \mathcal{C} \to \text{Sets} \) be a topological functor, and let \( D = \{ (C_1, T_1), \ldots, (C_n, T_n) \} \) denote a finite set of pairs \( (C_i, T_i) \) in which \( C_i \in \text{Ob}(\mathcal{C}) \) is an object and \( T_i \in J(C_i) \) is a trait classification on \( C \). The \textit{\( D \)-display of} \( F \) is

4. **Ontologies as sites, and sites as ontologies**

5. **Databases as ontologies**

6. **Morphisms and ontology integration**