Pragmatic-based Ontology Design and Alignment

Alexandra Mazak and Bernhard Wally

1 Vienna University of Technology
1 \{mazak, wally\}@big.tuwien.ac.at

Abstract. A multitude of research works and surveys dealing with ontology reuse point to a discrepancy between ontology engineers and users when interpreting the meaning of ontology concepts. Missing the engineers’ knowledge of the concepts’ usage in certain contexts may lead to problems (e.g., pragmatic heterogeneity) in the interpretation task caused by the lack of the users’ capability to comprehend design decisions (i.e., what they have in mind when describing a domain of interest). The intended use of concepts has a great impact on their interpretation. Therefore, we assume that meaning interpretation involves more than merely identifying the semantics of assumptions explicitly expressed by model theory-based techniques, since meaning depends on both semantics and pragmatics. For this purpose, we introduce a pragmatic-based approach for implementing an evidence-based (unidirectional) communication model from engineers to users, which is mainly influenced by the relevance-based inferential model of verbal communication.

1 Introduction

The vision of the Semantic Web as well as the Pragmatic Web is to provide a global infrastructure for the representation and exploitation of human knowledge. Ontologies are a central element of this vision providing the structural representation of that knowledge. They enable parties to communicate and exchange knowledge. Interoperability among ontologies is the major goal for realizing this vision. Therefore, ontologies, as resources of the web, have to be reconcilable for gaining both semantic and pragmatic interoperability.

Ontology alignment tools perform pair-wise comparison of entities from each of the source ontologies. Their algorithms are trying to find the best correspondences among these entities by selecting the most similar pairs. Ontologies are fairly complex structured, thus it is often practical to focus on different levels of ontologies separately, rather than trying to align them as a whole. The understanding of relations among entities of different ontologies is a cognitively difficult task, added with the user’s uncertainty about a possible mismatch risk resulting from heterogeneity factors among the sources. Often, this task is tedious, time-consuming, and error-prone, since it requires deep domain knowledge, which is not available to users. A major problem in ontology alignment is the lack of aids for supporting the interplay between the user, the tools, and that process.

Generally, ontologies are shared models and contexts are local models that encode the modeler’s subjective view of the domain. First of all, associated with ontology reuse, Giunchiglia [1] detects that context encodes an individual’s subjective view of the world, which is partial and approximate. In [2] the authors specify a “purely” subjective view of the domain as cognitive perspective. They state that this subtle form of perspective “is very important in the analysis of what is generally called an
intensional context” [2]. Obviously, there is a strong relation between the cognitive perspective, pragmatics, and contexts. However, the existing separation of contexts and ontologies leads to a “separation of meaning”, which constitutes an obstacle when providing users a single, consistent environment. Generally, semantics is expressed as context-independent meaning or sentence meaning, whereas pragmatics or cognitive semantics as context-dependent meaning. Fetzer [3] underpins this position. She points out that “meaning is at the heart of both semantics and pragmatics”; as a result of which we propose to consider pragmatics similarly to semantics at the ontology layer in order to fully support meaning consideration in the design process, as well as meaning interpretation in the alignment process.

2 Problem Description

Smart and Engelbrecht [4] present a small-scale pilot study with five participants. The aim of the study was to explore factors, which could determine why mismatches occur in the first place. The participants were asked to develop OWL\(^1\) ontologies describing the same domain of interest, a suicide bomb attack, at the same level of detail. The results reveal that no two subjects settle on the same representational solution, even though they were all provided with the same stock of information (e.g., concepts, relationship) about that specific domain. This fact leads the initiators to the conclusion that there are more profound (i.e., not only terminological) differences among entities. These are resulting from the differential use of ontology modeling formalisms to express content [4]. This means that ontology authors can use the same ontology language for describing an identical domain of interest, nevertheless, there are subtle differences among the domain ontologies. The study’s result brings forward that the cognitive state (i.e., mental state) of the ontology authors, rather than the semantics of the used language, causes these differences. The outcome of a previous study, analyzed in [5], comes to similar conclusions. In this study two ontologies explicitly represent two contexts, but they overlap on a common part. The authors point out that, despite of this overlap, there is no guarantee for users in ontology alignment that the modelers’ conceptualization of the common part is the same. The objections of the authors of both studies are confirmed by our own evaluation survey, which we conducted in the course of our dissertation [6]. There, the participants highly agree that the ontology authors’ intentional mental state on the domain, at design time, has significant impact on the usage of ontology entities (e.g., classes, relations).

These problems lead to pragmatic heterogeneity, which is mainly caused by differences in the modelers’ usage of entities in a certain (domain-related) context. That may lead to problems in the users’ interpretation. For instance: there are two classes identified as syntactically equal by an algorithm based on their label similarity

\[(O_A) : Author, (O_B) : author\]

a user cannot automatically infer that the classes’ meaning or relevance in use is similar, too; or that

\[(O_A) : Contribution, (O_B) : article\]

are similar, because they are used synonymously in the same context (e.g., to describe authors and their publications). Currently, this profound heterogeneity is not pre-

\(^1\) OWL = Web Ontology Language, http://www.w3.org/TR/owl-guide/ (last accessed January-03-2014).
dictable or solvable; neither by model theory-based methods nor by semiotic-based methods, which exploit the theory of signs. Therefore, pragmatic heterogeneity is still a continuous problem in ontology alignment. The main problem is to define usage patterns for discovering such a context-based similarity in an efficient way, since there is a strong relation between pragmatic and context.

3 Pragmatic-based Approach

The drawbacks posed in the previous chapter makes it necessary for us to follow a better integrated user support in ontology alignment that already starts when designing ontologies from scratch. Such an early consideration of alignment support is crucial in our approach. We present the idea to implement an evidence-based (unidirectional) communication model from engineers to users, which is mainly influenced by the relevance-based inferential model of verbal communication.

We view the ontology engineers’ design knowledge as the linkage of design rationales to design artifacts according to [7]. This knowledge captures formal as well as informal semantics. We assume that formal-based design decisions are related to the modeler’s logical perspective in compliance with an ontology language (e.g., OWL DL), whereas informal semantics are related to their pragmatic-based view of domain concepts. The logical view posits the kind of knowledge modelers have when they describe concepts well-formed by using the syntax and semantics of a language. For instance, that is knowledge of some necessary and sufficient defining conditions for category membership (e.g., anybody who is related to the class author is a type of person). Seen from this perspective an ontology is a set of assumptions about a domain explicitly expressed.

Semantic interoperability expresses the degree of correspondence between the set of expressions and the concepts of the domain to be modeled. An assumption expressed in OWL DL in the form of a statement is a certain item of information about a domain. For example: we want to express that papers are contributions.

We define two classes by declaring them as named classes by the labels: Paper and Contribution;

'Class(Paper)' ∈ r
'Class(Contribution)' ∈ r

in order to express that each paper is a contribution, we use the language construct rdfs:subClassOf;

'SubClassOf(Paper, Contribution)' ∈ r

The semantic representation of a statement deals with a sort of common meaning shared by every sentence of it. A well-formed sentence concerns the structure of a language, and therefore it has a fixed truth value with respect to that structure. From this logical perspective meaning is a function from assertions into truth values computed by deductive rules.

Our experience leads us to the assumption and the pilot study presented in [4] and a multitude of other works confirm it that design decisions are only to a small extent deductive driven. In the design process modelers have a comprehensive access to contextual information: intuition, belief, goal, experience, and others, which are cognitive abilities and as such pragmatic features [8]. We assume when making the
modeler’s pragmatic-based view of domain concepts visible to users that they would get a “bigger picture” of the sources as currently feasible. But this perspective is not exploitable by model theory.

3.1 Pragmatics and Relevance Theory

Pragmatics is the study of linguistic or speech acts and the context in which they are processed. “Context” is the all-pervasive concept of pragmatics [9]. There are two major theories of communication in pragmatics: the code and the inferential model. In the first theory, the focus is placed on a single model where communication is achieved by encoding and decoding messages. In the second theory, the communicator provides evidence of their intentions and the audience infers those intentions from that evidence. The evidence is a contextual information providing a precise purpose to the hearer. In the inferential model communication is achieved by producing and interpreting evidence based on the speaker’s informative communicative intention. We denote the representatives of the code model theory as speech act theorists, who take a classical view of pragmatics and those of the inferential model theory as relevance theorists, who focus on pragmatics interdisciplinary.

In our approach, we take a more relevance theoretic-based view of pragmatics, therefore we focus on the inferential and not the code model theory. Pragmatics from a relevance theoretic point of view is an investigation of “meaning in context” or “meaning in use”. The goal of the Relevance Theory, introduced by Sperber and Wilson [10], is to consider contextual effects of explicitly expressed assumptions by their relevance in the present context. In this theory relevance is an intuitive comparative and quantitative concept (e.g., irrelevant, weakly, relevant). Sperber and Wilson define relevance as a classificatory concept expressed in necessary and sufficient conditions;

“An assumption is relevant in a context if and only if it has some contextual effect in that context” [10].

The goal is to help focusing the attention of the audience on relevant information, because they are neither able to decode, nor to deduce the communicator’s intention of relevance.

4 Conceptual Design

We consider traditional ontology design and alignment as to be comparable with the classical pragmatic-based code model. Thereby, semantic interoperability is provided by model theory. We take a step forward by presenting the idea of implementing a system akin to the relevance-based inferential model in order to make it feasible for modelers to give information of the context-based relevance of the propositional formula to users. We consider implicitly made assumptions as informative contextual intentions that are based on the modeler’s pragmatic-based view. In our approach, modelers annotate the propositional formula with relevance-based meta-information when designing the ontology and users are able to interpret these meta-information before starting an alignment between two ontologies. In order to avoid misunderstandings we use the term “importance” instead of “relevance”. Based on the idea of weighted graphs, where semantic content can be communicated in the form of
weighting the nodes and edges, we implement an extended OWL DL metamodel by which the ontology’s relational structure can be enriched with pragmatic-specific instances.

Domain concepts and their relationships are in set theory classes and relations that constitutes the relational structure of the ontology. We focus on classes and their relations to other classes—together—as a subset, which can be annotated with importance weightings on that relations (cf. Figure 1). Such importance weightings give information of the entities’ usage in terms of their “relevance” in the domain description [11].

In OWL DL the rdfs:domain defines the kind of things the object property may apply to, whereas the rdfs:range defines the values the property is allowed to take. These axioms constitute logical constraints for semantic-based reasoning, while the importance weightings are pragmatic-based constraints and have as such no effects on model-based semantics. Figure 1 illustrates the approach. Each binary relation (owl:ObjectProperty) is annotated with an importance weighting depending on its domain/range axioms and on the relevance of that proposition in the domain contexts (i.e., its contextual effect).

**Definition 1. Importance weighting**
We define an importance weighted relation signature \( \sigma(R)_{iw} \) as a quadruple;

\[
\sigma(R)_{iw} = \langle \sigma, l_v, l_a, \omega \rangle
\]

where \( \sigma \) is the signature function; \( l_v \) is a function that maps the domain and range classes (i.e., nodes) to a set of natural language-based labels; \( l_a \) maps the binary relations to a set of arc labels; and \( \omega \) is a weighting function by which elements of \( \sigma \) can be annotated with elements of \( IW, \omega : (R \rightarrow C \times C) \rightarrow IW; IW \) is a finite set of importance weighting labels, \( IW = \{ \text{highest importance, high importance, middle importance, low importance, lowest importance} \} \).

For example;

\[
\omega(\sigma) : (\text{writes} \rightarrow (\text{Author, Contribution})) \rightarrow \text{highest importance}
\]

this means that the proposition “authors write contributions” is annotated by a pragmatic-based constraint in the form of an importance weighting label highest importance. The pragmatic-based condition “having a contextual effect to some degree” means in our approach that concepts are often classified as more or less
important concerning the fulfillment of the purpose-specific design goals of the ontology.

In order to make such a meta-annotation feasible, we define an extension to the OWL DL metamodel by adding Meta Object Facility [12] constructs for populating the ontology’s structure with pragmatic-specific instances. The language constructs of OWL DL that we use as input are the owl:ObjectProperty, the rdfs:domain and the rdfs:range axioms. The meaning of these basic elements from a logical perspective is already defined, which guarantees a shared interpretation. The kind of meta-annotation we propose is comparable to that of a propositional attitude in pragmatic theory, which constitutes the relation between a communicator and their uttered proposition (cf. Section 3).

5 Application in Ontology Alignment

Generally, the output of an alignment algorithm are lists of candidates. Tools which provide such lists for users are for instance: FOAM, Chimaera and PROMPT. Often, such lists are difficult for users to understand and interpret, e.g., due to poor readability. Mapping discovery is one of the major bottlenecks in ontology alignment. Currently, there is no support of filtering large lists in order to categorize or group candidates by certain characteristics. Therefore, it is difficult for users to get a quick and context-based overview of the sources, much less to know which concepts are the core concepts and which ones are good candidates as initial points, e.g., for schema-based alignment techniques.

By using the introduced approach domain concepts can be ranked by their importance weighting. This makes it feasible for users to detect at a glance the concepts’ information significance in the domain context, which makes it easy to identify all core concepts. For instance, users can be guided when outlining an interesting sub-scene of the sources, e.g., for candidate selection or to define initial points. We implement an algorithm for grouping concepts in lists sorted in descending order based on their importance weightings (cf. Table 1).

The rankings in Table 1 are based on pragmatic based view of the participants of our evaluation survey [6]. The scores are computed as average of the values resulting from the importance weighting procedure. For instance, users can easily detect that the core concepts of the two ontologies are: Author of O_A and author of O_B, which are also syntactically equal; and Contribution of O_A and article of O_B. Additionally, the lists may help users to take care of terminological heterogeneity, which occurs due to variations in names referring to the same concepts, like in case of Contribution/article. This means that both classes might be used to describe the same thing—a written contribution to a conference. The two terms are used synonymously, but it is not straightforward to detect them as similar, neither by string-based techniques nor by users if they are not aware of the domains’ context. Moreover, users can detect that the intended usage of the classes is quite similar. Thus, they can infer that aligning O_A and O_B carries a minimum risk of pragmatic heterogeneity in itself.

6 Conclusion

The main issues of our approach of a pragmatic-based ontology design and alignment are: (i) we view model-theoretic semantics as the relation between the logical form
Table 1. Grouped classes of two example ontologies by their importance weightings.

<table>
<thead>
<tr>
<th>Level</th>
<th>confOf ((O_A))</th>
<th>crs_dr ((O_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest</td>
<td>Contribution</td>
<td>article</td>
</tr>
<tr>
<td>Author</td>
<td></td>
<td>author</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>abstract</td>
</tr>
<tr>
<td>Middle</td>
<td></td>
<td>reviewer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>review</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest</td>
<td>Administrative_event</td>
<td>conference</td>
</tr>
<tr>
<td></td>
<td>Working_event</td>
<td>program</td>
</tr>
<tr>
<td></td>
<td>Organization</td>
<td>chair</td>
</tr>
<tr>
<td></td>
<td>Person</td>
<td>participant</td>
</tr>
<tr>
<td></td>
<td>Member_PC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scholar</td>
<td></td>
</tr>
</tbody>
</table>

of expressed statements (sentence meaning) and the entities in the real world; and (ii) cognitive semantics as the relation between that statements and their meaning in a certain context (sentence meaning in context), as intended by the original modelers; (iii) we adapt pragmatics viz. its relevance theory in order to make cognitive semantics visible to users when aligning ontologies. We stated that it is important to distinguish between the modelers’ logical view in combination with the used ontology language, and their pragmatic-based view of the domain. The latter subjective view facilitates the analysis of the intensional context (cf. Section 1). For its consideration we presented the idea of a relevance-theoretic system that interfaces with the use-conditional system. For this purpose, we adapted the relevance-based inferential communication model of ostensive behavior (cf. Section 3.1). Our aim is to improve model-based reasoning methods by an additional importance-driven evidence-based decoding system and not to replace them.

References


