Towards ontology interoperability through conceptual groundings

How to cite:

For guidance on citations see FAQs

© 2008 The Authors
Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Towards Ontology Interoperability through Conceptual Groundings

Stefan Dietze, John Domingue
Knowledge Media Institute,
The Open University,
MK7 6AA, Milton Keynes, UK
{s.dietze, j.b.domingue}@open.ac.uk

Abstract. The widespread use of ontologies raises the need to resolve heterogeneities between distinct conceptualisations in order to support interoperability. The aim of ontology mapping is, to establish formal relations between a set of knowledge entities which represent the same or a similar meaning in distinct ontologies. Whereas the symbolic approach of established SW representation standards – based on first-order logic and syllogistic reasoning – does not implicitly represent similarity relationships, the ontology mapping task strongly relies on identifying semantic similarities. However, while concept representations across distinct ontologies hardly equal another, manually or even semi-automatically identifying similarity relationships is costly. Conceptual Spaces (CS) enable the representation of concepts as vector spaces which implicitly carry similarity information. But CS provide neither an implicit representational mechanism nor a means to represent arbitrary relations between concepts or instances. In order to overcome these issues, we propose a hybrid knowledge representation approach which extends first-order logic ontologies with a conceptual grounding through a set of CS-based representations. Consequently, semantic similarity between instances – represented as members in CS – is indicated by means of distance metrics. Hence, automatic similarity-detection between instances across distinct ontologies is supported in order to facilitate ontology mapping.

Keywords: Semantic Web, Ontology Mapping, Conceptual Spaces, Interoperability.

1 Introduction

The widespread use of ontologies - formal specifications of shared conceptualisations [10] - together with the increasing availability of representations of overlapping domains of interest, raises the need to resolve heterogeneities [12][14] by completely or partially mapping between different ontologies. With respect to [2][17], we define ontology mapping as the process of defining formal relations between knowledge entities which represent the same or a similar semantic meaning in distinct ontologies [6][19]. In that, ontology mapping strongly relies on identifying similarities [1] between entities across different ontologies. However, with respect to this goal, several issues have to be taken into account. The symbolic approach - i.e. describing symbols by using other symbols, without a grounding in the real world - of established representation standards such as OWL¹ or RDF-S² which are based on first-order logic (FOL) and syllogistic reasoning [8] leads to ambiguity issues and

¹ http://www.w3.org/OWL/
² http://www.w3.org/RDFS/
does not entail meaningfulness, since meaning requires both the definition of a
terminology in terms of a logical structure (using symbols) and grounding of symbols
to a conceptual level [3][16]. Therefore, concept representations across distinct
ontologies – even those representing the same real-world entities - hardly equal
another, since similarity is not an implicit notion carried within ontological
representations. But manual or semi-automatic identification of similarity
relationships – based on linguistic or structural similarities across ontologies
[13][7][9] – is costly. Consequently, representational facilities, enabling to implicitly
describe similarities across ontologies are required in order to support ontology
interoperability.

**Conceptual Spaces (CS)** [8] follow a theory of describing entities at the conceptual
level in terms of their natural characteristics similar to natural human cognition in
order to avoid the symbol grounding issue [3][16]. In that, CS consider the
representation of concepts as vector spaces which are defined through a set of quality
dimensions. Describing instances as vectors enables the automatic calculation of their
semantic similarity by means of spatial distance metrics, in contrast to the costly
representation of similarities through symbolic representations. However, several
issues still have to be considered when applying CS. For instance, CS do not
explicitly prescribe any applicable representation method. Moreover, CS provide no
means to represent arbitrary relations between concepts or instances, such as part-of
relations. In order to overcome the issues introduced above, we propose a two-fold
knowledge representation approach which extends FOL ontologies with a conceptual
grounding by refining individual symbolic concept representations as particular CS.
Consequently, similarity becomes an implicit notion of the representation itself,
instead of relying on manual or semi-automatic similarity detection approaches.

2 Conceptual Groundings for Ontological Concepts

With respect to the aforementioned issues, we argue that basing knowledge models on
just one theory alone might not be sufficient. Therefore, we propose a two-fold
representational approach – combining FOL ontologies with corresponding
representations based on CS – to enable similarity-based reasoning across ontologies.
In that, we consider the representation of a set of \( n \) concepts \( C \) of an ontology \( O \)
through a set of \( n \) Conceptual Spaces CS. Hence, instances of concepts are
represented as members in the respective CS. While still benefiting from implicit
similarity information within a CS, our hybrid approach allows overcoming CS-
related issues by maintaining the advantages of FOL-based knowledge
representations. In order to be able to represent ontological concepts within CS, we
formalised the CS model into an ontology, represented through OCML [15]. Hence, a
CS can simply be instantiated in order to represent a particular concept.

Referring to [8][18], we formalise a CS as a vector space defined through quality
dimensions \( d_i \) of CS. Each dimension is associated with a certain metric scale, e.g.
ratio, interval or ordinal scale. To reflect the impact of a specific quality dimension on
the entire CS, we consider a prominence value \( p \) for each dimension [8]. Therefore, a
CS is defined by \( CS^e = \{ p_1 d_1, p_2 d_2, \ldots, p_n d_n \} | \forall d_i \in CS, p_i \in P \} \), where \( P \) is the set of real
numbers. However, the usage context, purpose and domain of a particular CS strongly
influence the ranking of its quality dimensions what supports our position of
describing distinct CS explicitly for individual concepts. Please note that we do not distinguish between dimensions and domains [8] but enable dimensions to be detailed further in terms of subspaces. Hence, a dimension within one space may be defined through another CS by using further dimensions [18]. In this way, a CS may be composed of several subspaces, and consequently, the description granularity can be refined gradually. Dimensions may be correlated. Information about correlation is expressed through axioms related to a specific quality dimension instance.

A particular member \( M \) – representing a particular instance – in the CS is described through valued dimension vectors \( v_i \) like \( M = [v_1, v_2, ..., v_n] \). With respect to [18], we define the semantic similarity between two members of a space as a function of the Euclidean distance between the points representing each of the members. However, we would like to point out that different distance metrics, such as the Taxicab or Manhattan distance [11], could be considered, dependent on the nature and purpose of the CS. Given a CS definition \( CS \) and two members \( V \) and \( U \), defined by vectors \( v_0, v_1, ..., v_n \) and \( u_0, u_2, ..., u_n \) within \( CS \), the distance between \( V \) and \( U \) can be calculated as

\[
\text{dist}(u, v) = \sum_{i=1}^{n} p_i \left( \frac{(\bar{u} - \bar{v})}{s_u} - \frac{(v_i - \bar{v})}{s_v} \right)^2 \]

where \( \bar{u} \) is the mean of a dataset \( U \) and \( s_u \) is the standard deviation from \( U \). The formula above already considers the so-called Z-transformation or standardization [4] which facilitates the standardization of distinct measurement scales in order to enable the calculation of distances in a multi-dimensional and multi-metric space.

**Representing Ontological Concepts through Conceptual Spaces**

The derivation of an appropriate space \( CS_i \) to represent a particular concept \( C_i \) of a given ontology \( O \) is understood a non-trivial task which primarily implies the creation of a CS instance which most appropriately represents the real-world entity represented by the symbolic concept representation. We foresee a transformation procedure consisting of the following steps:

S1. Representing concept properties \( pc_{ij} \) of \( C_i \) as dimensions \( d_{ij} \) of \( CS_i \).
S2. Assignment of metrics to each quality dimension \( d_{ij} \).
S3. Assignment of prominence values \( p_{ij} \) to each quality dimension \( d_{ij} \).
S4. Representing instances \( I_{ik} \) of \( C_i \) as members in \( CS_i \).

A specific CS is instantiated by applying a transformation function which is aimed at instantiating all elements of a CS (S1 – S3). S1 aims at representing each concept property \( pc_{ij} \) of \( C_i \) as a particular dimension instance \( d_{ij} \) together with a corresponding prominence \( p_{ij} \) of a resulting space \( CS_i \):

\[
\text{trans} : \{ pc_{i1}, pc_{i2}, ..., pc_{in} \} \rightarrow \{ p_{i1}d_{i1}, p_{i2}d_{i2}, ..., p_{in}d_{in} \} \forall d_{ij} \in CS, p_{ij} \in P
\]

Please note that we particularly distinguish between data type properties and relations. While the latter represent relations between concepts, these are not represented as dimensions since such dimensions would refer to a range of concepts (instances) instead of quantified metrics, as required by S2. In the case of relations, we propose to maintain the relationships represented within the original ontology \( O \) without representing these within the resulting \( CS_i \). In that, the complexity of \( CS_i \) is reduced to
enable the maintainability of the spatial distance as an appropriate similarity measure. 

$S2$ aims at the assignment of metric scales (interval scale, ratio scale, nominal scale), while $S3$ is aimed at assigning a prominence value $p_{ij}$ to each dimension $d_{ij}$. Prominence values should be chosen from a predefined value range, such as $0...1$. With respect to $S4$, one has to represent all instances $I_k$ of a concept $C_i$ as member instances in the created space $CS_i$. This is achieved by transforming all instantiated properties $pi_{ikl}$ of $I_{ik}$ as valued vectors in $CS_i$. 

$$\text{trans}: \{p_{i_{a1}}, p_{i_{a2}}, ..., p_{i_{ae}}\} \{I_{ik} \in PL_i \} \Rightarrow \{v_{a1}, v_{a2}, ..., v_{ae}\} \{I_{ik} \in M_a\}$$

Hence, given a particular CS, representing instances as members becomes just a matter of assigning specific measurements to the dimensions of the CS. In order to represent all concepts $C_i$ of a given ontology $O$, the transformation function consisting of the steps $SI-S4$ has to be repeated iteratively for all $C_i$ which are element of $O$. The accomplishment of the proposed procedure results in a set of CS instances which each refine a particular concept together with a set of member instances which each refine a particular instance.

3 Conclusion

In order to facilitate ontology mapping, we proposed a hybrid representation approach based on a combination of FOL ontologies and multiple concept representations in individual CS. Representing concepts following the CS theory enables representation of instances as vectors in a respective CS and consequently, the automatic computation of similarities by means of spatial distances. A CS-based representation is supported through a dedicated CS formalisation, i.e. a CS ontology, and a formal method on how to derive CS representations for individual concepts. Within proof-of-concept prototype applications, e.g. [5], an OCML [15] representation of the proposed hybrid representational model was utilized to validate the applicability of the approach. Following our two-fold representational approach supports implicit representation of similarities across heterogeneous ontologies, and consequently, provides a means to facilitate ontology mapping. Moreover, our approach overcomes certain individual issues posed by each of the two approaches. Whereas traditional ontology mapping methodologies rely on mechanisms to semi-automatically detect similarities at the concept and the instance level, our approach just requires a common agreement at the concept level since similarity information at the instance level is implicitly defined.

However, the authors are aware that our approach requires a considerable amount of additional effort to establish CS-based representations. Future work has to investigate this effort in order to further evaluate the potential contribution of the approach proposed here. Moreover, further issues related to CS-based knowledge representations still remain. For instance, whereas defining instances, i.e. vectors, within a given CS appears to be a straightforward process, the definition of the CS itself is not trivial at all and dependent on subjective perspectives. With regard to this, CS do not fully solve the symbol grounding issue but to shift it from the process of describing instances to the definition of a CS. Nevertheless, distance calculation relies on the fact that resources are described in equivalent (or mapped) geometrical spaces. However, we would like to point out that the increasing usage of upper level ontologies and the progressive reuse of ontologies, particularly in loosely coupled
organisational environments, leads to an increased sharing of ontologies at the concept level. As a result, our proposed hybrid representational model becomes increasingly applicable by further enabling similarity-computation at the instance-level towards the vision of interoperable ontologies.

4 References