Evaluating the Structural Quality of Semantic Hierarchy Alignments

Cliff Joslyn
Knowledge Systems Group
Pacific Northwest National
Laboratory
Richland, WA, 99352, USA
cjoslyn@pnl.gov

Alex Donaldson
National Security Directorate
Pacific Northwest National
Laboratory
Richland, WA, 99352, USA
alex.donaldson@pnl.gov

Patrick Paulson
Knowledge Systems Group
Pacific Northwest National
Laboratory
Richland, WA, 99352, USA
patrick.paulson@pnl.gov

ABSTRACT

We present preliminary results of applying a novel method based in metric order theory to provide a measure of the structural quality of some of the test alignments between semantic hierarchies used in the 2007 Ontology Alignment Evaluation Initiative.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Semantic networks

Ontology alignment [2] is increasingly seen as a critical Semantic Web technology. There is currently a great effort in alignment in application areas like computational biology, where top-down, monolithic development of unitary ontologies is at best difficult, and at worst undesirable [5, 7].

While there is a diversity of approaches and tools to aligning ontologies (e.g. [6]), to the extent that ontologies are dominated by their taxonomic cores (that is, of their subsumptive inheritance and/or meronomic compositional class hierarchies), then aligning such semantic hierarchies specifically becomes a central task for ontology matching in general.

Our approach [3, 4] is to use order and lattice theory [1] to model semantic hierarchies as finite, bounded, partially ordered sets (posets) $\mathcal{P} = \langle P, \leq \rangle$, so that nodes $a \in P$ are ontology concepts related by `is-a` or `part-of` links through the relation $\leq$. Note that such structures are not, in general, trees, or even lattices, but can be rich in multiple inheritance and lack unique least common subsumers between nodes.

Ontology alignment can then be cast as analyzing mappings $f\colon \mathcal{P} \to \mathcal{P}'$ or relations $F \subseteq P \times P'$ which take concepts in one semantic hierarchy $\mathcal{P} = \langle P, \leq \rangle$ into another $\mathcal{P}' = \langle P', \leq' \rangle$. We call such functions f or relations F **alignments** between **anchors** $a \in P, a' \in P'$, and we then desire to consider the question whether for a fixed pair of hierarchies $\mathcal{P}, \mathcal{P}'$, one such f or F is better than another, and in particular whether we can measure the structural properties of such mappings to help determine this.

Consider the example in Fig. 1 showing a mapping relation F between two posets $\mathcal{P}, \mathcal{P}'$. Then two independent criteria for the structural quality of F are:

Order Preservation: F should not distort the order of concepts. For example, F takes E and B, which are comparable in $\mathcal{P}$, to I and J, which are non-comparable in $\mathcal{P}'$.

Distance Preservation: F should not distort the metric relations of concepts. For example F takes both B and E, which are somewhat distant in $\mathcal{P}$, to the single node I in $\mathcal{P}'$, so that there is no distance between them on the right.
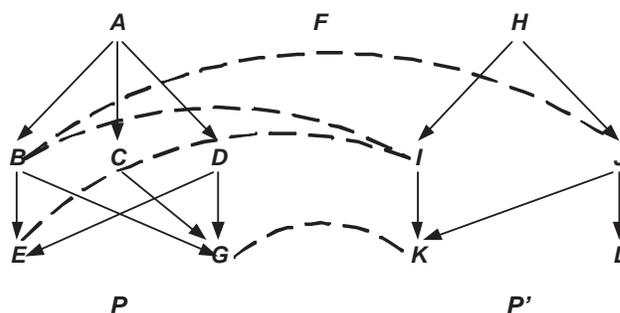

Figure 1: Example semantic hierarchy alignment relation.

In this poster we present some preliminary results from applying a novel method based in metric order theory to provide such a structural analysis of a portion of the test ontology alignments used in the 2007 Ontology Alignment Evaluation Initiative (OAEI[1]). The structure of the OAEI is to identify target ontologies to align, develop by hand a reference alignment, and then test submitted alignments based on precision and recall of matched nodes compared to that reference. In contrast, our objective is to develop absolute measures of the structural quality of any mappings between pairs of semantic hierarchies.

We developed and tested our method against the "food" track of OAEI 2007[2], which targets alignments between the UN Food and Agriculture Organization AGROVOC thesaurus (28K multilingual terms) and the US National Agricultural Library Agricultural Thesaurus (NALT, 42K English terms) using relations from the SKOS Mapping Vocabulary[3]. This track was selected because it was the only one

---

[1]http://oaei.ontologymatching.org/2007
[2]http://oaei.ontologymatching.org/2007/food
[3]http://www.w3.org/2004/02/skos/mapping/spec

with moderate size, relatively rich multiple inheritance, and a publicly available reference alignment. Within the track, there were three relations from the SKOS Mapping Vocabulary supported, but we focused only on "exactMatch", to emulate the general problem as we have modeled it.
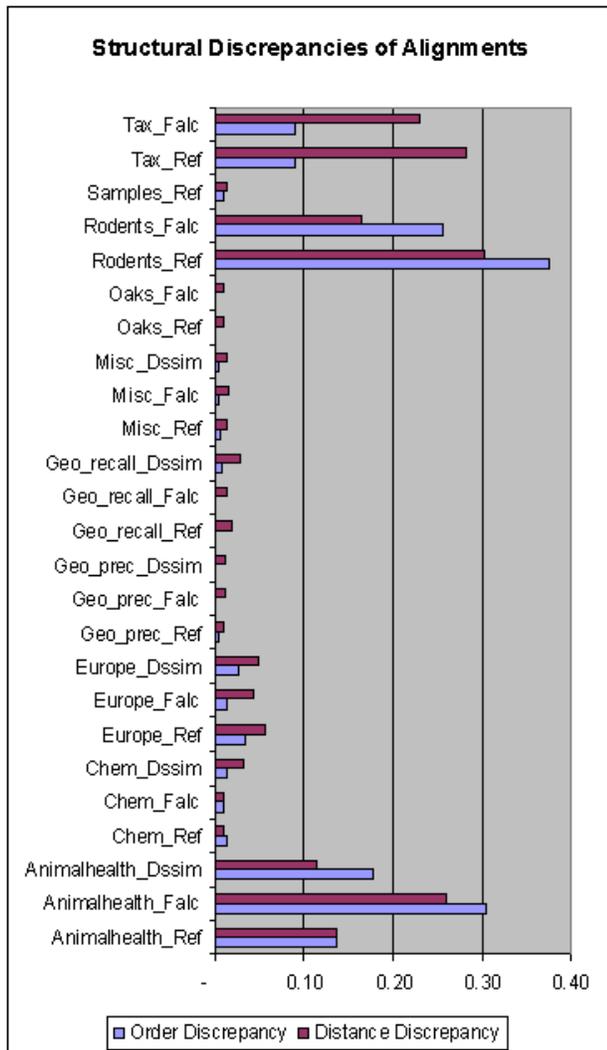
Figure 2: Measures of structural quality between reference and submitted alignments from the Food track of OAEI 2007.

Not only are these alignments allowed to take nodes on either side to multiple nodes on the other, but they do so in large numbers. This made the alignment relation representation F more appropriate than the alignment function f. Then, the reference alignment for this task consisted of a collection of ten separate `rdf` files, each one identifying a different subject matter such as "rodents" or "animal health". We measured the overlap amongst these groups and found them to be generally small, so that they are largely disjoint from each other. We thus considered these groups as separate alignments, and collected the anchors $\{a \in P : \exists a' \in P', \langle a, a' \rangle \in F\} \subseteq P$ and $\{a' \in P' : \exists a \in P, \langle a, a' \rangle \in F\} \subseteq P'$ linked on each side of each alignment, together with their ancestors and descendants, into collections called **facets**. Submitted alignments were not

broken out according to these facets, but rather extended across the breadth of the NALT and AGROVOC ontologies. To compute similar facets for the submitted alignments, we identified and collected submitted nodes matching reference nodes on each side, and their mappings, as the anchor sets.

Fig. 2 shows some of our early results. Alignments are identified by their names, and suffixed with either `Ref` for reference or the code of a submitter. Order discrepancy is a measure of the amount of order violations induced by F, while distance discrepancy is a measure of the difference of distances between pairs of nodes on each side. Methodological details will be provided in the full poster.

Preliminary considerations indicate a general correlation between order and distance discrepancies, but some substantial differences as well; subsantial variation in their overall magnitude between groups; and substantial agreement between reference and submitted alignments. Our continuing work is to analyze these results with respect to such considerations as the conditions leading to differences in order and distance discrepancies; the dependence with variation in sizes of the structures on the left and right, and amount of multiple inhertiance present; and potential visualization approaches.

1. REFERENCES

[1] Davey, BA and Priestly, HA: (1990) *Introduction to Lattices and Order*, Cambridge UP, 2nd Edition

[2] Euzenat, Jérôme and Shvaiko, P: (2007) *Ontology Matching*, Springer-Verlag, Hiedelberg

[3] Joslyn, Cliff: (2004) "Poset Ontologies and Concept Lattices as Semantic Hierarchies", in: *Conceptual Structures at Work, Lecture Notes in Artificial Intelligence*, v. **3127**, ed. Wolff, Pfeiffer and Delugach, pp. 287-302, Springer-Verlag, Berlin

[4] CA Joslyn, SM Mniszewski, A Fulmer, and G Heaton: (2004) "The Gene Ontology Categorizer", *Bioinformatics*, v. **20**:s1, pp. 169-177

[5] Kirsten, Toralf; Thor, Andreas; and Rahm, Erhard: (2007) "Instance-Based Matching of Large Life Science Ontologies", in: *DILS 2007, Lecture Notes in Bioinformatics*, v. **4544**, ed. S Cohen-Boulakia and V Tannen, pp. 172-187, Springer-Verlag, Heidelberg

[6] Noy, Natasha and Musan, Mark A: (2003) "The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping", *Int. J. Human-Computer Studies*, v. **59**, pp. 983-1024

[7] S Zhang, P Mork, O Bodenreider, PA Bernstein: (2007) "Comparing Two Approaches for Aligning Representations of Anatomy", *AI in Medicine*, v. **39**, pp. 227-236