Measures of similarity for integrating conceptual geographical knowledge: some ideas and some questions

Mark Gahegan, Ritesh Agrawal, Anuj Jaiswal, Junyan Luo and Kean-Huat Soon

GeoVISTA Center, Department of Geography, The Pennsylvania State University, USA
mark@geog.psu.edu

Introduction

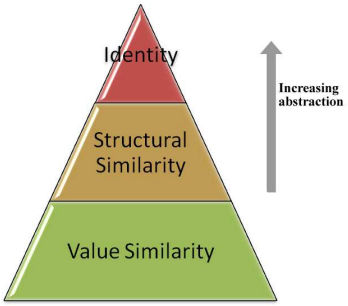
Knowledge computing has opened the possibility of representing and reasoning with knowledge about the geographical world, which (just like with data before it) has in turn led to many further questions regarding interoperability, integration and conflation of such knowledge.

In our research at the GeoVISTA Center at Penn State, we often find ourselves faced with the problem of reconciling knowledge captured from different experts, or mined from published documents and databases. Far from being a frustration, this work—for us at least—represents the intellectual heart of geography: i.e. finding ways to represent and share our understandings of Earth's systems. Such demands have led us to construct various ontology matching methods that we describe below, in Section 1. These methods are implemented in *ConceptVista*, a concept mapping, semantic search and knowledge integration environment. (http://www.geovista.psu.edu/ConceptVISTA/index.jsp).

However, there are also a number of broader questions about geographical knowledge integration that require very careful attention, and they relate to the philosophy of Ontology (big O), via the fields of hermeneutics, pragmatics and situated cognition. These questions arise, we believe, because of the situated and often contested sense of meaning that is common within geography and the natural sciences. The concepts and relations we use to describe the world do not exist in nature, they are *entirely* constructed by humans. Hence it is not surprising that meaning differs between individuals, and through time. Thus there are no perfect solutions for this kind of knowledge integration, but rather subjective measures and practices that, on balance, provide useful results. We introduce some of these issues in Section 2.

Section 1: Similarity measures for geographical knowledge integration.

In line with the object-oriented notions of equivalence based: (i) object identity, (ii) object structure (e.g. types of its properties and relations) and (iii) their actual values (instance similarity), similarity can be computed between ontological fragments using such measures (or by summation between whole ontologies).



Identity

In the general case, using object identity as a means to recognize two concepts as being equivalent requires assumptions regarding some over-arching or perhaps universal ontology from

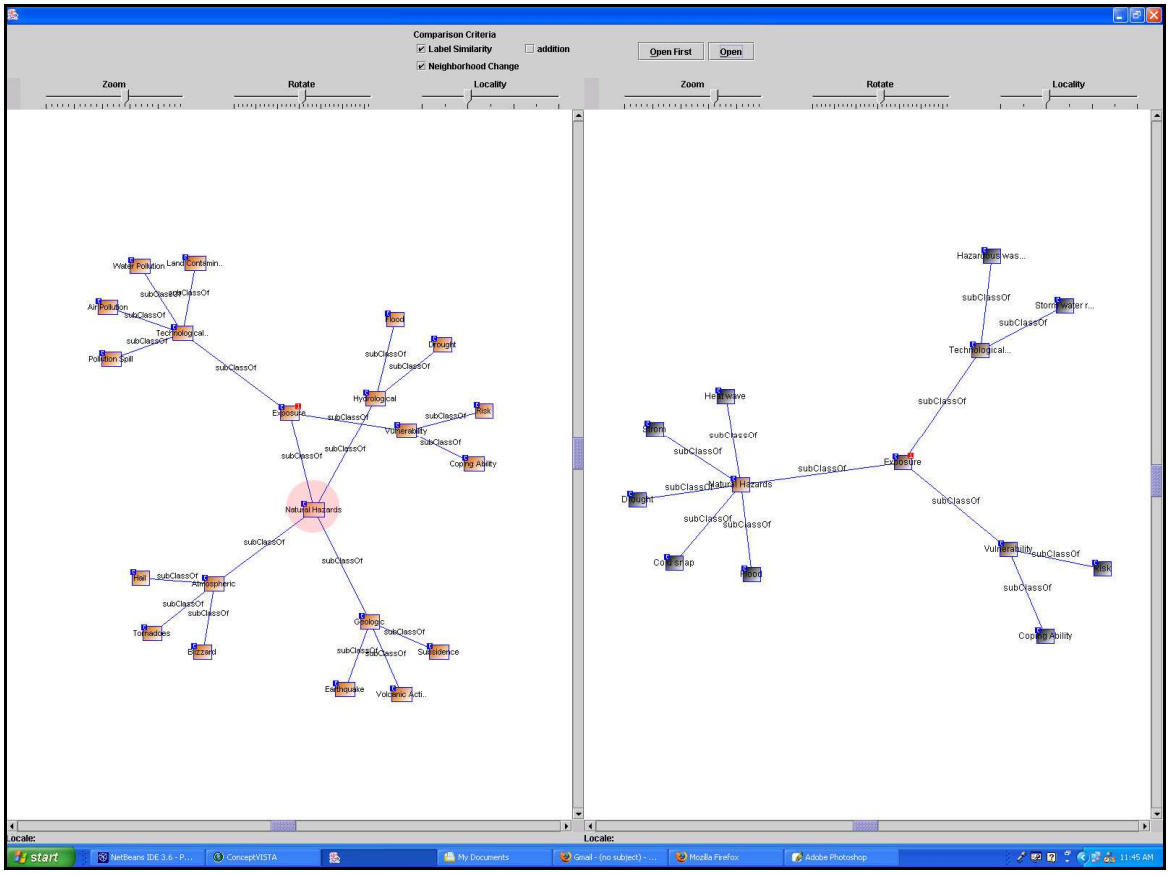
which both concepts are drawn. Again, in the general case, such ontologies do not exist as yet (and it is an interesting question as to whether they ever can/should exist in an open system such as the Earth). Where concepts are drawn from the same ontology, recognition of equivalence is trivial; in our work, each concept is assigned its own URI as a unique identifier that we use to recognize equivalences.

Structural similarity

There are many structural measures that can be used to calculate a similarity measure, for example:

- Number of properties common to two concepts
  This drives at the definitional structure of concepts, how they are understood
- Number of relationships common to two concepts
  This seems to relate more strongly to the role concepts play, how they are used
- The converse, number not of properties and relationships not common
- The total number of properties and relationships used by two concepts (their union); to normalize the result.

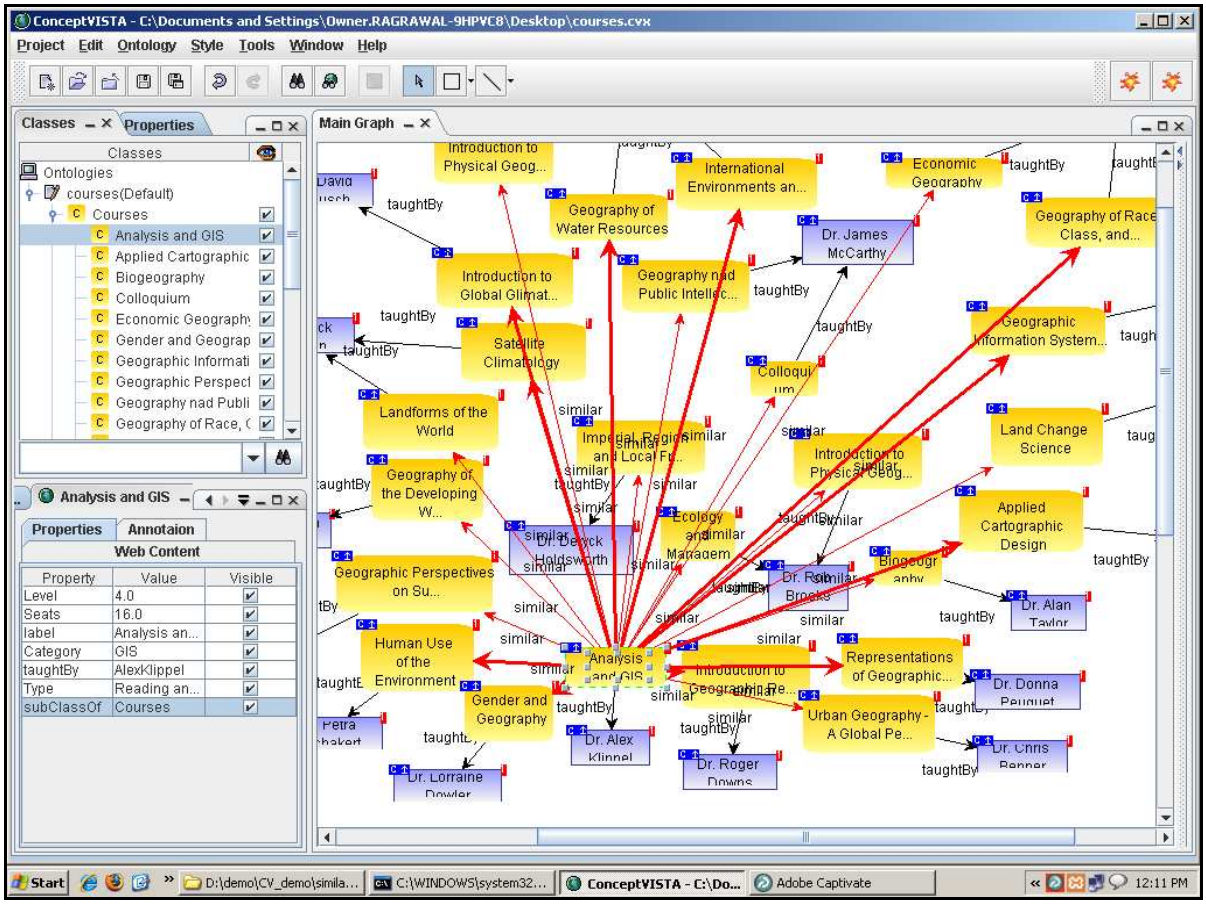
Various combination of such measures are helpful in identifying possible candidates for matching concepts that might appear in different ontologies, and are not sensitive to differences in names assigned to concepts, which is often useful. We have found these kinds of measures do help in recognizing similarities between semi-formal knowledge (e.g. concept maps) as the figure below shows.



Here, the problem is to try to extract common structure from concept maps (captured from geographers). These concept maps describe different opinions on the most important factors that have changed the landscape of certain geographical places. On the left side, a concept is highlighted; concepts on the right side are then colored according to the closeness of their match—based on their structural similarities.

Instance similarity (similarity by value)

Comparing the values of concept properties has also proved useful to us, as the figure below shows. Methods used here are similar to those described above for structural similarity. The figures shows Geography Courses offered at Penn State (yellow rectangles) and their Instructors (blue rectangles). Similarity between courses is shown as the thickness of the red line connecting the course currently in focus (*Spatial Analysis*) to all other courses.



Section 2: Broader questions relating to semantic similarity

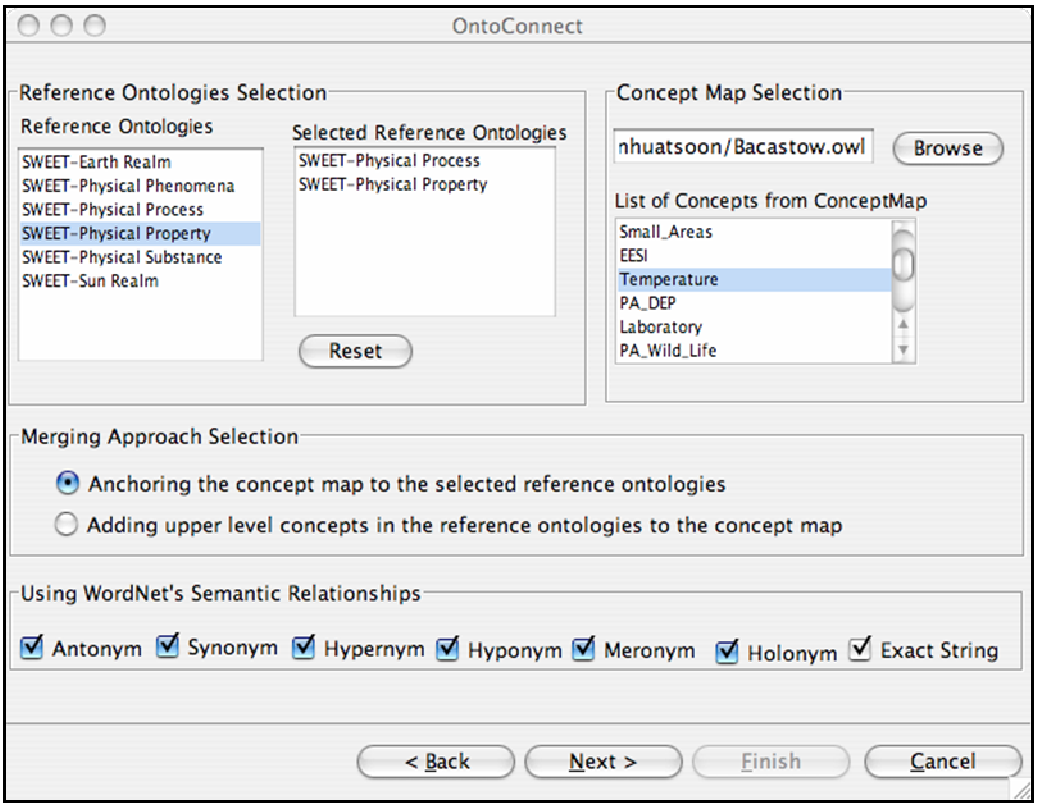
Semantic relationships connecting concepts

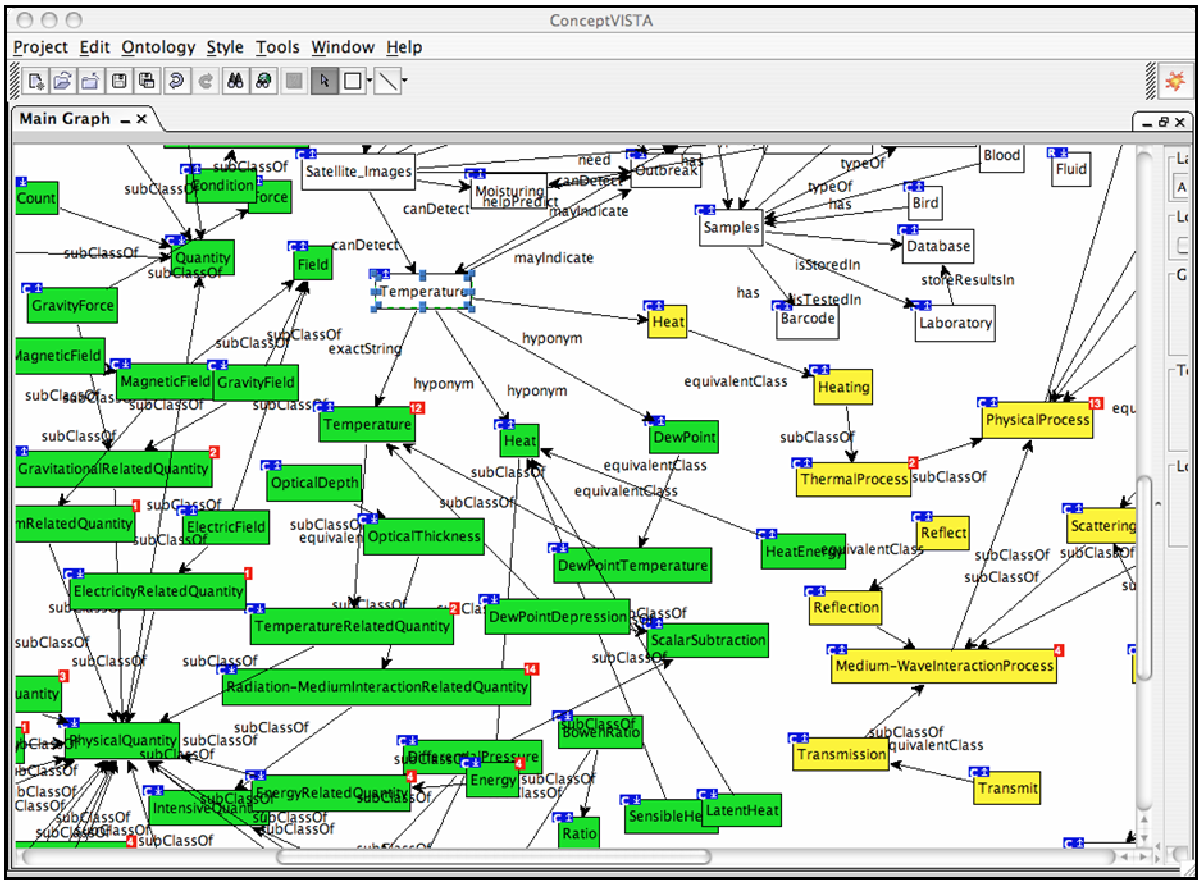
The above examples in Section 1 only consider the problem of recognizing that two concepts are similar and therefore might stand for the same idea. However, it is often the case when combining knowledge gathered from different human experts that we need to instead recognize

that concepts are connected by certain semantic relationships (specifically: Toponyms, Hypernyms, Hyponyms, Meronyms, Holonyms, Synonyms, and Antonyms). This poses some much harder problems: although Wordnet, Cyc and other thesauri and gazetteers can help, there are a multitude of such relationships that link many concepts together.

To try to address this problem we have developed an experimental semantic matching tool that enables us to merge domain ontologies with more informal knowledge (concept maps) using the semantic relationships from WordNet (e.g. hyponym, meronym). The tool searches for concepts that are similar in the concept map and domain ontologies based on these semantic relationships. The respective ontologies that have similar concepts are then merged with the concept map to form a conflated ontology. This leads naturally to the question of exactly how such merging should take place. Should properties from matching concepts be merged into one new concept? Should they be subsumed into one or more existing concepts? When combining knowledge across two communities, concepts—though their names and other properties may be similar—may actually mean different things to their knowledge community, or play significantly different roles. Thus assuming they are equivalent (and can therefore be merged) may be unhelpful.

The two figures below show the experimental system we have constructed to try and find answers to some of these questions. The panel on the left shows how various concept matching strategies can be initialized, and the panel on the right shows a result of merging concepts from an informal concept map into the SWEET EarthRealm ontology





Some Deeper Questions that motivate our interest in attending the workshop

1. **Which semantic relationships are the most useful in representing similarity, or in merging ontologies?** This remains an open question that we are very interested in, and probably requires some sophisticated experimentation with users to better understand with tools such as the one shown above. We are keen to hear of experiences from other researchers in this area.

2. **Do experts from different domains really refer to the same concept?** To each it may mean something different. This begs the further question: whose frame of reference does any merged ontology actually represent? It is certainly possible that concepts might act as boundary objects between two knowledge communities, but in this case should we recognize them as such, somehow?

3. **How should all these different measures of similarity be combined when they are used?** Is it ever possible to achieve useful matching results without supervision by human experts? Are there contexts in which some measures work better than others, and if so, what are they?

4. **Is there a useful role for hermeneutics to play in constructing knowledge horizons or some other form of perspective around concept maps, and showing where two horizons**

might intersect? We have experimented with local perceptual filters that can be applied in concept maps and ontologies in this regard but again we would like to hear from other researchers, particularly those looking into the problems of differences between individuals in interpreting conceptual knowledge.