A Unified Ontological Framework for Semantic Integration

Marinos Kavouras
OntoGEO Research Group
National Technical University of Athens
Outline

- Context - Problem - Objective
- Ontology types
- The Framework
- Approaches
Context I

- Research on “ontologies”, “integration” and “semantic interoperability”.

- Ontological approaches have been accepted as a very promising (if not the only) approach to semantic interoperability.

- Geographic reality (beings and kinds) and the way such knowledge is (or, better, should be) represented in modern information systems.

- Such a progress however, in addition to solving the intended problems, it has to resolve confusion caused by the different approaches developed by an interdisciplinary community for a variety of problems and domains.
Context II

Notions used differently.

TWO PERSPECTIVES

A “higher” ontological perspective

domain and conceptualization differences

A “lower” design/implementation perspective

explication differences

The “ontology of ontologies”
Existent geographic ontologies are developed according to different semantic contexts.

Differences in the conceptualization and categorization of geographic concepts may not be easily identified and formalized due to their weak semantic content.

Therefore, they raise problems when ontologies from heterogeneous contexts are to be integrated.

Most available approaches to semantic integration provide ad-hoc, non-systematic, subjective manual mappings (in the best case), often leading to procrustean amalgamations (in the worst), in order to fit the target standard.
Objective

- Evaluation and comparison of the approaches is very difficult and impedes usability as well as methodological interoperability.

- In order to facilitate understanding, this paper presented the key issues of semantic integration in a unified framework.
Ontology types

Ontologies are classified along two dimensions:

- the level of formality, and
- the level of granularity or generality.

According to the level of formality, three ontology types are specified:

- An informal ontology is the simplest type; it comprises of a set of concept names organized in a hierarchy.
- A terminological ontology consists of a hierarchy of concepts defined by natural language definitions.
- A formal ontology further includes axioms and definitions stated in a formal language.
According to the **level of granularity**, ontologies are classified into six types:

- **top-level** - very general concepts such as space, time, object, event, etc., independent of a particular domain.
- **general** - a large number of concepts related to fundamental human knowledge.
- **domain** - concepts associated with a specific domain
- **task** - concepts related to the execution of a particular task or activity.
- **application** concepts essential to a particular application.
- **meta-ontology** or **generic** or **core** ontology - concepts which are common across various domains; can be further specialized to domain-specific concepts.

With respect to ontology association, we can distinguish the input ontologies often called **resource** ontologies, and the possibility of using common ontologies to assist/coordinate the mapping called **shared ontologies** or **target ontologies**.
Framework

- 3 processes on the road to semantic integration
- collaboration of the processes
- still distinct having their own merit
- their principal characteristics are introduced below...

Fig. __: Basic processes towards semantic integration.
Principal Characteristics

- Assumptions made
- Semantic level addressed
- Input (source) / Output components
- Method used
- Degree of change
- Degree of interaction

Secondary characteristics:

- **knowledge representation paradigms** (Description/Frame-based logic)
- **conceptual structures used** (Conceptual Graphs, Concept Lattices, etc.)
- **ontology languages** (RDFS, OIL DAML+OIL, OWL, etc.)
Process I: Semantic Information Extraction

- **Source components:** free text, corpora, thesauri, specialized text (e.g., definitions), terms, nomenclatures, data dictionaries, hierarchical classifications, database schemata, etc.

- **Taxonomic ontologies vs. axiomatized or formal ontologies.**

- **What constitutes semantic information?**

- **Empirical ad hoc approaches attempting to formalize the concepts involved, and design the associated databases.**

- **Information extraction (IE) approaches based on NLU/NLP - central terms in computational linguistics and artificial intelligence.**
(...cont.) Process I

- Reveal salient semantic information in geographic ontologies (properties and relations) (Kokla & Kavouras, 2002; Tomai & Kavouras, 2003).
Process II: Concept/Ontology comparison

- Comparison and similarity measures reveal/depict how difficult integration (Process III) will be.

- A comparison shall reveal and somehow measure similarities or heterogeneities (conflicts).

- Similarity between geographic concepts can be estimated by combining feature and linguistic matching, and semantic distance calculation (Tversky, 1977; Rodríguez & Egehofer 2002; Yaolin et al. 2002).

- Process II needs also to resolve the heterogeneities.
Semantic heterogeneities between different ontologies are caused by:

- **different coverage (level of detail)** due to different scope - user needs,

- **different relations** due to imposition of single inheritance or due to different classification perspectives,

- **different semantics** due to different conceptualizations - classification perspectives.
Conceptualization of a real world entity consists of a term $T$ and a definition $D$ (e.g., natural language definition) - a concept $C = (T, D)$.

- Clearest cases: “equivalence” (same term and definition) and “disjointness” (different terms and definitions).
- “Synonymy” occurs when two concepts are represented by different terms (e.g., “pond” and “pool”) with the same definition (“small lake”).
- “Overlap” occurs when the terms are the same (e.g., “canal”) but the definitions overlap (e.g., “manmade or improved natural waterway used for irrigation and transportation” and “artificial waterway used for recreation and transportation”).
Process III: Integration

- **Alignment** is a light mapping between concepts of different ontologies. No ontology is distorted. Translation/conversion utilities provide functionality.

- **Partial compatibility** causes a unification of those parts of ontologies that are considered more similar. The rest of the parts are still necessary. The unified parts have distorted the initial common ontology parts.

- **Unification** extends partial compatibility to all ontologies and their concepts. Each resource ontology is forced/distorted to become the same (thus fully compatible) with the others. A single ontology to deal at the end. Resource ontologies are not anymore usable.

- **True integration** creates a single integrated ontology whose parts are the resource ontologies including some additional concepts necessary for the association. The user deals with a single integrated ontology. Resource ontologies are not distorted retaining their independence and usability.
Alignment

Partial compatibility

Unification

True integration
(...cont.) Process III

Integration Methods are distinguished with respect to:

**D1.** The possible **change/ alteration/ distortion** caused. Two cases: **NO/ YES** (0/1).

**D2.** The **number** of ontologies resulting. Two cases: **SINGLE/ MANY** (0/1).

**D3.** The use of a **target/ shared ontology** in the process. Two cases: **NO/ YES** (0/1).

<table>
<thead>
<tr>
<th>#</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Integration</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Integration targeted</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Alignment</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Alignment targeted</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Unification</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Unification targeted</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>Partial compatibility</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Partial compatibility targeted</td>
</tr>
</tbody>
</table>

**Classification of ontology integration approaches.**
Approaches

- Various research projects: MOMIS, KRAFT, PROMPT, Chimaera and ODEMerge, etc.

- A survey on ontology tools by Ontoweb, 2002.

- In the geospatial domain, there have been an increasing number of approaches to semantic interoperability (Visser et al., 2000; Kavouras & Kokla, 2002; Fonseca et al., 2003) addressing different issues and focusing on diverse application needs.

- Table demonstrates how the unified framework presented above, based on three processes and their principal characteristics, helps understand and evaluate the integration approaches.
<table>
<thead>
<tr>
<th>Method</th>
<th>Source components</th>
<th>Interaction</th>
<th>Method</th>
<th>Components used</th>
<th>Heterogeneities identified / resolved</th>
<th>Interaction</th>
<th>Method</th>
<th>Change caused</th>
<th>Num.of Result onto/s</th>
<th>Use Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MOMIS</strong></td>
<td></td>
<td>semi-automatic</td>
<td>Genera-</td>
<td>terminologic</td>
<td>ARTEMIS (tool based on affinity-based clustering</td>
<td></td>
<td>semi-automatic</td>
<td>no</td>
<td>many</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>tion of a common thesaurus</td>
<td></td>
<td>tion of a common thesaurus</td>
<td>intensional and extensional relationships</td>
<td>(tool based on affinity-based clustering techniques)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(plan for the future)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Schema-/lexicon</td>
<td>Names + attributes</td>
<td>Semantically related classes are clustered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>derived relationships, (WordNet), designer-</td>
<td></td>
<td>together</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>supplies relationships, inferred relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KRAFT</strong></td>
<td></td>
<td>semi-automatic</td>
<td>NLP technique s + connectio-</td>
<td>Text material</td>
<td>Identify heterogeneities between resource ontologies + ontology mappings between resource ontologies and shared ontology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n to a top-level ontology</td>
<td>(probably) manually</td>
<td>Find semantically similar entities, then compare attributes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Terms and attributes</td>
<td>Classification of ontology mismatches according to Visser et al., 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PROMPT</strong></td>
<td></td>
<td>Semi-automatic</td>
<td>-</td>
<td>-</td>
<td>Linguistic similarity (any -term matching algorithm) + clues about ontology structure and user actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Names of classes and slots, class hierarchy, slot attachment to classes, facets and facet values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OntoGEO</strong></td>
<td>Minimal user involve-</td>
<td>Minimal user involve-</td>
<td>-</td>
<td>-</td>
<td>Connection to on-line lexicon to compare synonymous values of the same semantic info</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ment.</td>
<td>ment.</td>
<td></td>
<td></td>
<td>Terms, semantic relations and properties, hierarchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NLP technique s to extract semantic info from definitions</td>
<td>Definitions of terms</td>
<td></td>
<td></td>
<td>Semantic heterogeneities</td>
<td>Minimal user involve-</td>
<td>Construction of a concept lattice using FCA</td>
<td>no</td>
<td>one</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

- This paper presented the **key issues of semantic integration** in a **unified framework**.

- This framework assists the user in **making the right decisions in an integration endeavor**.

- Further **research** is needed in all aspects of the **three processes** involved.

- Various **ontological issues**:
  - extraction of semantic relations and properties,
  - integration of geographic ontologies,
  - association to top-level and basic-level ontologies,
  - semantic similarity,
  - ontological vagueness,
  - context/scale and granularity issues,
  - guidelines to ontology generation.

- [http://ontogeo.ntua.gr/](http://ontogeo.ntua.gr/)