# Creating a Geospatial and Visual Information Ontology for Analysts

Chumki Basu[1], Hui Cheng[1], Christiane Fellbaum[2]
[1]Sarnoff Corporation, 201 Washington Road, Princeton, NJ, 08543
[2]Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544

## Abstract

*An ontology is a main component of an evolving knowledge base that caters to multiple clients. Consider a scenario where an automated procedure (a computer vision algorithm) used in an analyst tool detects different kinds of "roads" in images, and features in the ontology are used to distinguish a "paved" road from a "dirt road". In another scenario, the ontology enables reasoning about "locations", supporting analysts' geospatial information processing tasks. In this paper, we describe the creation of a multi-use geospatial and visual information ontology, GVIO[1], building on and integrating with the lexical database, WordNet. To ensure that GVIO can interoperate with other ontologies in useful ways, we inherit as much of the WordNet structure and content as is relevant for the domain of aerial surveillance and link in new content/structure as necessary.*

## 1. Introduction

Geospatial and visual information are essential to intelligence gathering. There is a need to associate meaning with the kinds of entities and relationships useful for information processing tasks (e.g., geospatial query of a region) [1]. In this paper, we describe a Geospatial and Visual Information Ontology (*GVIO*) we are developing for analyst-specific information processing tasks and computer vision applications. This is a challenging problem as the requirements for these applications can be very different. Whereas an analyst may be interested in locating "facilities near *CityX*", the input requirement of a vision algorithm (used in an analyst tool) could be salient "characteristics" of buildings in and around *CityX*.

## 2. Problem approach

We are interested in understanding how an analyst analyzes the content of aerial video and imagery. There is no single source of knowledge that sufficiently characterizes the information necessary for this type of analysis. Instead, there are a variety of independent resources including *WordNet* [2], *GML*[2] (Geography Markup Language), *LSCOM* [3], *Cyc* [4], and subject matter experts (SMEs).

We start with the lexical database, *WordNet*, as our semantic base. Similar to Swartout *et al*. [5], we create an ontology using top-down and bottom-up methods. Our goal is to capture a mix of high, mid-level and domain-specific terms in the ontology, while maintaining the distinction between types and instances defined in *WordNet.*

### 2.1. Top-down *WordNet* filtering

We filtered top-level categories in *WordNet* (Table 1), pruning concepts that need not be further examined (e.g., *Cognition*, *Food*, *Feeling* and *Motivation*). We manually classified categories as geospatially/visually relevant, neither, or mixed (relevant and non-relevant). Some categories are mixed and may not be pruned significantly (less than 25%). Other categories (e.g.,

[1] This work was supported by DTO/ODNI under the CASE program. We thank Dan Doney, Emile Morse, and Dennis Moellman for useful discussions. We also thank Glenn Petry at Sarnoff Corporation for his contributions to the reasoning application.

[2] http://www.opengeospatial.org/standards/gml

*Phenomenon*, *Causal Agent*) need further inspection to determine the amount to prune (labeled, "undetermined"). Since the distribution of terms across top-level categories is not uniform (e.g., *Event* has a large number of hyponyms), we were left with many unexamined nodes.

### 2.2. Bottom-up data collection

We generated an analyst survey of 400+ terms distilling analyst searches for aerial video/satellite imagery into three lexical categories: *nouns*, *verbs* and *adjectives*. This survey includes an SME concept list for 2/3D computer vision object detection tasks in the urban environment. We list sample terms from each lexical category in Table 2.

We link (map) the terms to *WordNet* synsets. This is a manual step due to polysemy – e.g., we disambiguate the intended sense of "apron", "a paved surface where aircraft stand while not being used" (ruling out the "protective garment" reading of this word). Rank ordering the terms by respective hyponym tree sizes, we list the top-10 terms in descending order (1). The result is a significant reduction in the total number of relevant or mixed synsets – less than 25% of the total number of synsets in *WordNet.* Combining with top-down filtering, we achieve further pruning (e.g., of terms appearing in the top-level *Person* category).

```
Person, Location, Tree, Move,(1)
Leader, Vehicle, Water, Ground,
Building, Grass
```

### 2.3. Defining properties

We defined a set of properties for each lexical category. Visual properties of a physical entity are features useful for object detection [6][7]. To maximize utility for object detection algorithms, properties should be quantified, if possible – i.e., assigned default values or ranges of values. For example, we know "telephone pole", an artifact, has some average "height" based on instances of telephone poles observed. Properties also have

| Category | Filtering result |
|---|---|
| *Location* | <25% |
| *Event* | 0% |
| *Act* | <25% |
| *Artifact* | <25% |
| *Phenomenon* | Undetermined |
| *Entity* | <25% |
| *Attribute* | <25% |
| *Measure* | <25% |
| *Cognition* | 100% |
| *State* | Undetermined |
| *Time* | 0% |
| *Substance* | >75% |
| *Relation* | >75% |
| *Person* | >75% |
| *Communication* | >75% |
| *Causal Agent* | Undetermined |
| *Possession* | Undetermined |
| *Group* | <25% |
| *Food* | 100% |
| *Shape* | 0% |
| *Natural object* | <25% |
| *Feeling* | 100% |
| *Animal* | >75% |
| *Plant* | >75% |
| *Motivation* | 100% |

**Table1: Top-down *WordNet* filtering**

| Nouns | Verbs | Adjectives |
|---|---|---|
| airfield | carry | armored |
| barn | chase | barren |
| hospital | enter/exit | civilian |
| loading dock | load/unload | dark |
| telephone pole | meet | rocky |

**Table2: Sample terms in analyst survey**

associated subsumption hierarchies – e.g., in Figure 1, "height" is specialized as "sitting height" and "standing height" (useful for pose detection) and default values are assigned for "male" and "female" (derived from anthropometric studies[3]). In Figure 2, hyponyms of "car"

[3]http://ergo.human.cornell.edu/DEA325notes/anthrodesign.html

inherit “dimension” properties. From *WordNet*, the subsumption hierarchy for “dimension” includes properties, “height”, “width”, and “length”. *CityGML*[4] defines properties for urban settings (e.g., buildings); we link these properties to our ontology, as appropriate.

## 3. Spatial Reasoning from Text

An interesting analyst application that uses ontological relationships is reasoning about spatial entities in text to search imagery/video. Simple keyword-based search is prone to vocabulary mismatches in query terms vs. index terms (from annotations). Often, the annotation space is sparse, resulting in missing data (e.g., location names). Consequently, search queries with location keywords will return no results. A key challenge is the following:

New sites will not be labeled in imagery/video. How do we retrieve imagery/video that contain these locations?

We transform a text-based query (2) into a spatial query expressed in geo-coordinates (latitude, longitude) in multiple steps (Figure 3).

```
The ABC Training Center is    (2)
20 kilometers northeast of
CityX.
```

Using a named-entity detector[5], we find location terms in the original text. We disambiguate a missed detection – “location” incorrectly labeled as “organization” or “person” – using *WordNet*. Using a predefined set of *WordNet* location categories (`{WN-Locations}`), including “city”, “state”, “country”, “capital”, “lake”, “river”, “building”, etc., and the “instance of” and “hypernym” relationships, we define a function, `inferLocation`:

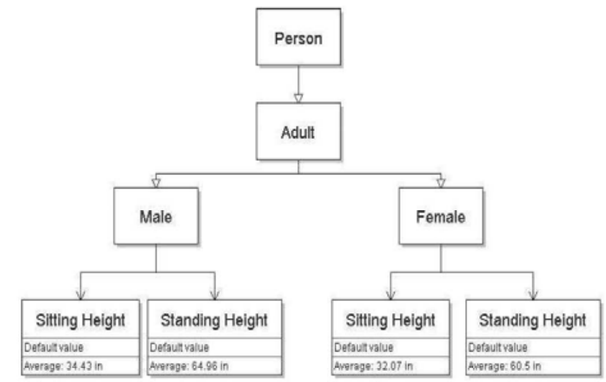


**Figure 1: Visual property, “height”, specialized as “sitting height” and “standing height” for concept, “person”**

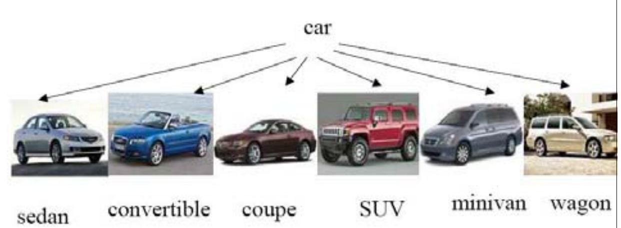


**Figure 2: Types of “car” inherit “dimensions” (“height”, “width”, “length”) as visual properties**

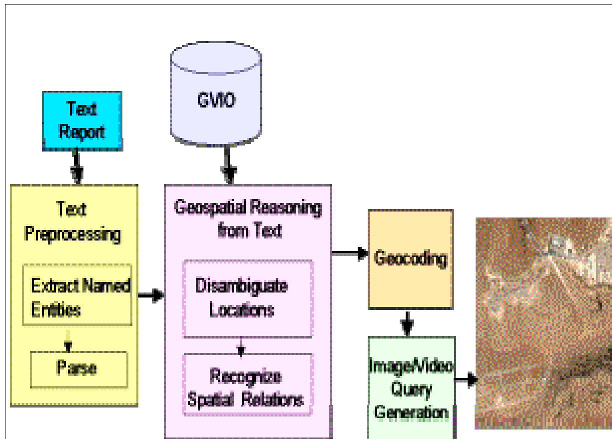


**Figure 3: Flow Diagram**

4 http://www.opengeospatial.org/standards/gml
5 http://alias-i.com/lingpipe/

| | |
|---|---|
| Direction/Distance | **North/South/East/West of**<br>Far **from, Near** |
| Quantifier + preposition | 20 miles **From/West of,**<br>Very **Near**/Far **from** |
| Simple prepositions | **In**, **On**, **At, …** |

**Table 3. Prepositions for analysis**

| |
|---|
| **Configuration1**: Location1 is (located, found) Rel Location2 |
| **Configuration2:** Location1 and Location2 are Rel (near, far, south of each other, …) |
| **Configuration3:** There is Location1 Rel Location2. |
| **Configuration4:** Location1 is Rel (south of, far, near, …) Location2 |
| **Configuration5:** Ellipsis: Only Location1 is mentioned, Location2 is implied. |

**Table 4. Spatial configurations in text**

```
inferLocation(entity) =
1 if instanceOf(entity)∈
  {WN-Locations} or
 hypernym(entity)∈{WN-Locations},
0 otherwise
```

Prepositions are highly polysemous, which makes disambiguating meaning very challenging [8][9][10]. Table 3 provides a partial list of prepositions/relations to be analyzed. We choose syntactic configurations from the list in Table 4 to disambiguate spatial readings between two locations.

## 4. Conclusions

In this paper, we presented a multi-use geospatial and visual information ontology. We described how object detection algorithms and a geospatial reasoning application benefit from ontology content. We continue to develop this ontology into a general-purpose resource that can be used by analysts.

## References


[1] M. J. Egenhofer. Towards the Semantic Geospatial Web. *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, McClean, Virginia, November, 2002.

[2] C. Fellbaum. *WordNet: An Electronic Lexical Database*, MIT Press, 1998.

[3] M. R. Naphade, A. Hauptmann, S. Chang, and J. R. Smith. A Large Scale Concept Ontology for Multimedia Understanding, Technical Report, 2007.

[4] D. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, 1990.

[5] B. Swartout, R. Patil, K. Knight, and T. Russ. Toward Distributed Use of Large-Scale Ontologies, *Proceeding of the Tenth Knowledge Acquisition for Knowledge-based Systems Workshop*, November 9-14, 1996, Banff, Canada.

[6] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer. Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2., Madison, WI, 2003.

[7] R. Srihari and Z. Zhang. Show & Tell: a Semi-automated Image Annotation System. *IEEE Multimedia*, 7(3):61-71, July 2000.

[8] P. Saint-Dizier (ed.), *Syntax and Semantics of Prepositions*, Springer, 2006.

[9] W. G. Hayward and M. J. Tarr. Spatial Language and Spatial Representation. *Cognition 55*, 39-84, 1995.

[10] A. Herskovits. Language and Spatial Cognition: An Interdisciplinary Study of the Preposition in English, *Studies in Natural Language Processing*, Cambridge University Press, 1986.