The Role of *Place* for the Spatial Referencing of Heritage Data

Krzysztof Janowicz
janowicz@uni-muenster.de
Institute for Geoinformatics, University of Münster, Germany

Abstract

The workshop on *The Cultural Heritage of Historic European Cities and Public Participatory GIS* aims at outlining the challenges to develop a ppGIS which can integrate data from different ages, authorities, and formats. This data ranges from archival catalogs, maps, photographs, film, aural archives over archaeological data to interpretative studies. With respect to spatial resolution, data should be available down to the level of single streets or even specific properties. Instead of a static approach, users should also be able to contribute new data to the system. This raises a couple of questions which are also relevant in a broader context, namely the ontological representation of and reasoning about geographic places. This paper introduces the challenges of modeling places in such a heterogeneous setting, outlines steps towards a three layered solution on how to represent, reference, and reason about geographic places, and finally sketches how to deal with inconsistent and contradictory knowledge.

1 Ontological Perspectives on the Notion of Place

Geographic places are abstract entities used to structure knowledge and to ease communication. Usually they refer to a physical region (i.e., an extent) in space and are categorized according to some commonly agreed upon characteristics. As social entities, places are of interest for individual communities in a certain region and for a particular time span (see [5, 2] for an overview). Names are a common handle to refer to places [9]. However, places can also be referred to by ostention or (narrative) descriptions. Consequently, the same place can be referred to by various names by different people at different times or even by placeholders such as *Anyshire*. Similarly, the location and spatial extent the place refers to may be vague, change over time, or even be unknown. While places are abstract we can still experience them by a number of perceivable characteristics of the physical region they refer to [3, 12]. Examples are the surface and texture of the physical body of the earth, artifacts such as buildings, as well as knowledge extracted from information carriers such as maps or photographs. Beyond directly perceivable characteristics, places may also be classified by convention, e.g., as administrative areas.

The clear separation of the physical region and the social construct called *place* also allows to model phenomena such as places which disappear or move without running into logical contradictions. A temporal normand settlement,

moves when the perceivable characteristics move instead of the region on the earth's surface [12]. A place disappears when all its carriers disappear, e.g., if all persons aware of the place are no longer alive and all artificial carriers, e.g., maps, are lost or destroyed. Nevertheless, a disappearing place has no influence on the physical region it refers to. Finally, a place name such as USSR can be out-dated in a sense that it still refers to the same physical region but is only used in a historical context.

While the notion of place is central for many knowledge organization systems, it is difficult to formulate rigid identity criteria for places. As minimum definition, gazetteers consider named geographic places as triples of names, spatial footprints, and types (N, F, T) [9]. However, gazetteers such as the Alexandria Digital Library Gazetteer and the Getty Thesaurus of Geographic Names also take spatial containment, administrative parthood, and temporal scopes into account [10]. This allows to state that the populated place called York was part of the Roman Empire and is now part of the UK. Nevertheless, this is not sufficient to answer the question of place identity and especially not whether and why two names refer to the same place. As mental handles places do not depend on specific names, geographic feature types, nor a stable and well defined portion of space. Similarly, a place does not lose its identity if administrative areas and higher order divisions change. The city of York, for instance, was renamed several times, had different spatial extents, changed its type from fortress to city over the centuries, was home of different cultures, different religions and languages, but we still refer to it as one consistent and continuous place.

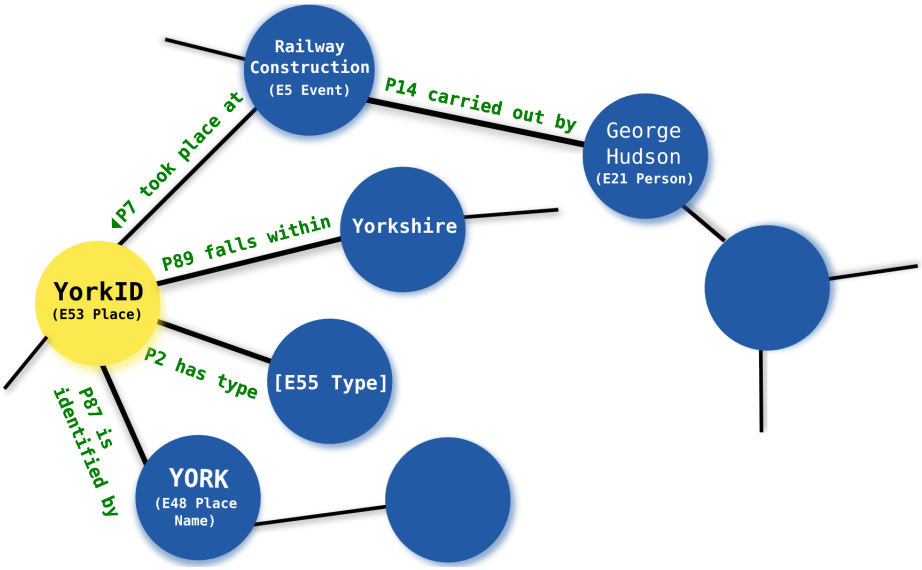

Figure 1: Modeling knowledge about the city of York using CIDOC CRM [6].

Figure 1 shows some historical facts about the city of York organized using CIDOC CRM [6]. CIDOC CRM is a well established and standardized (ISO 21127) Conceptual Reference Model (CRM) developed and maintained by the Committee on Documentation (CIDOC). It is intended as a top-level ontology for the annotation of heterogeneous cultural heritage data to make it available

in a machine-readable format (RDF) for knowledge extraction, integration, mediation and interchange [6]. The graph shows a RDF-resource called *YorkID* as RDF-subject as well as several predicates and RDF-objects which are subjects of other RDF-triples again. The CIDOC CRM predicate *P87 is identified by* is used to link the resource *YorkID* to another resource of type *E48 Place Name* (which is, same as *E47 Spatial Coordinates*, a subtype of *E44 Place Appellation*) that represents the literal York. The place YorkID has a particular type, e.g., *City*, and falls within another place – in this case the administrative area called Yorkshire. One strength of CIDOC CRM is its focus on actions and relationships instead of a detailed subtype hierarchy. For instance, we can state that a particular event (*E5 Event*) took place at the city of York, namely the construction of the railway. The arrow in front of the predicate indicates that it is the inverse relation of *P7b witnessed*. This event was carried out by a Person (*E21 Person*) called George Hudson and so forth.

Note that for reasons of readability, the graph in Figure 1 contains many simplifications, for instance, the separation between the resource and its name (appellation) is left aside for Yorkshire and George Hudson. Similarly, the railway construction should be linked to a certain period such as the time of industrialization, and so forth. As CIDOC CRM is a top-level ontology it does not define particular geographic feature types or types of artifacts. Instead *E55 Type* acts as a proxy to domain and application level ontologies. The type *City* in Figure 1 may be taken from a Feature Type Ontology (FTO) for geographic places [12].

While CIDOC CRM was created to describe data from the domain of cultural heritage, resources are not restricted to paper documents, maps, photographs, and so forth, but can also refer to real world entities such as specific exhibits in a museum. To account for the vagueness and incompleteness of historical data, CIDOC CRM does not define mandatory predicates for most types and, besides domains and ranges, also does not implement constraints for their usage. For instance, it is not necessary to define a birth event (*E67 Birth*) for each person, and the birth event may link to two different places if the annotated historical sources disagree in this point.

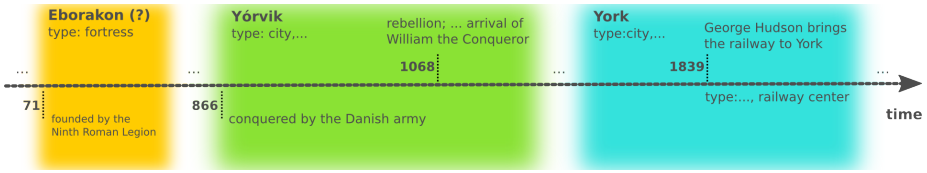

Figure 2: A time-line for the place York.

A weakness of the place model implicitly underlying the York example in Figure 1 is its temporal scope. In fact, it is atemporal. The place named York today had other names before, and while it is a city nowadays it was created as a fortress as part of the Roman Empire. Consequently, many of the RDF-triples require a limited temporal scope within which they are valid. Figure 2 reflects this view by introducing a time line for York. For reasons of simplification, we distinguish time frames by place name here, however, of course names may change stepwise over time and do neither correspond to crisp boundaries nor does a change in the naming always indicate a new temporal scope (time frame).

As depicted in Figure 2, York became a railway center in the 19th century which was not the case before. Similarly, it was part of several empires in different time frames and so forth. One could also represent that places split, move, or merged over time [15]. To take a different example, two countries merged 1989/90 in the German Reunification and consequently Germany has different capitols in different time frames.

The views depicted in Figure 1 and 2 do not contradict and one can map between them in principle[1]. Nevertheless, they focus on different aspects of the notion of place. While in the first case, place is a nexus connecting Persistent Items (E77) to Temporal Entities (E2), the second view focuses on modeling places by their time-lines and in fact equates *place* with its history. In this view, there is a place which was founded by the Romain Empire, later conquered by the Danish army, etc, and is referred to by the literal York these days. This also changes the perspective on place identity; a place remains the same if we can (in principle) draw a continuous time-line for it. While there are also counter examples for such simplified view, one could argue that places can be understood as perdurants; we experience them indirectly by the participation of endurants[2]. These endurants can be actors or physical items in general and may again participate in other perdurants such as events. For instance we can directly experience the arrival of a certain person, e.g., William the Conqueror, but not the rebellion and conquest. These events can again only be perceived by changes they cause. Modeling places as perdurants also goes along with recent findings from cognitive science about artifacts in general. The HIPE theory of function [1] for instance explains the function and categorization of artifacts by their **H**istory, **I**ntentional perspective, the **P**hysical environment, and **E**vent sequences. An affordance-based and therefore user-centric approach to the notion of place has been presented by Jordan et al. [14].

Finally, a last example pointing out that the notion of place is not restricted to a spatial extent on the surface of the earth is the place where Horatio Nelson died – the deck of the H.M.S. Victory. In this case, the place is part of a man-made object (E22); see [6] for details. While the deck exists before Nelson's death the place is only relevant in context of this event. Moreover, as long as we assign the place to the deck of the ship and not its location during the naval battle at Trafalgar, the place moves in space whenever the H.M.S. Victory moves. This has some surprising consequences on spatio-temporal reasoning and especially spatial containment. From this view point, the place of Nelson's death is in Portsmouth, England. A promising approach to explain why the deck of a ship can be a place in its own rights is to model places as thematic roles [19]. Being a place is a role entities can play if they are used to reference other entities (which could also be events) in space. Such a role is only played temporally as demonstrated in case of the deck of the H.M.S. Victory.

One important aspect of thematic roles is their dependency on perspective. The fact that William the Conqueror arrived at York may be modeled in two conceptual graphs [19] using different thematic roles:

[1] However, one has to keep in mind that CIDOC CRM was created for monotonic reasoning.

[2] 'Classically, endurants (also called continuants) are characterized as entities that are 'in time', they are 'wholly' present (all their proper parts are present) at any time of their existence. On the other hand, perdurants (also called occurrents) are entities that 'happen in time', they extend in time by accumulating different 'temporal parts', so that, at any time t at which they exist, only their temporal parts at t are present.'[18, p.11]

(a) [Person: William] ← (Agent) ← [Arrive] → (Location) → [York]
(b) [Person: William] ← (Agent) ← [Go] → (Destination) → [York]

According to Sowa's hierarchy of thematic roles [19], *Location* is a subrole of *Essence* (which again is a subrole of *Product*). In contrast, *Destination* is a subrole of *Goal* (which is also a subrole of *Product*). Other spatial roles are *Origin* and *Path*. Consequently, York can play the role of a location, destination, or origin in different situations but also in the same situation described from different points of view. Note that thematic roles should not be mixed up with subsumption relations, i.e., York plays the role of a location but is not of type *Location* in an ontological sense.

While these are initial thoughts for a conceptualization of place applicable for the envisioned ppGIS, the next section demonstrates how thematic data can be used as additional reference to disambiguate places.

2 Thematic Data as Additional Reference

Our knowledge about historical places is often vague, incomplete, or even contradictory. Even in the same administrative area place names are not necessarily unique, change over time, may refer to prominent landmarks (such as a telegraph station) which do not exist any more, and so forth. Several authorities and historical documents may refer to the same place by different names and report on the events which took place there from many different (and contradictory) perspectives. Nevertheless, we want to structure heterogeneous data stored in several formats and make it available in a consistent way using a ppGIS. Consequently, it is necessary to group data which refer to the same place and separate it from data which, in fact, refer to another place. One approach would be a semi-automatic service to generate identity assumptions about historical places [11]. As depicted in Figure 3, the core idea of such an identity assumption service is to use the links established (e.g., by CIDOC CRM annotations) between places, actors, objects, and events as additional points of reference.

In its broadest definition, geoinformation consists of a spatial, a temporal, and an attributive (also called thematic) component [7]. The interpretation of the first two components is given by well defined spatial and temporal reference systems such as coordinate systems and calendars. In analogy, the third component requires attributive or semantic reference systems [4, 17]. While spatial reference systems can be used to measure distance in space, semantic reference systems can be used to measure the semantic distance, called similarity [13], between types and individuals.

We argue that place names referred to in historical documents probably refer to the identical (real world) place if they are related through the same or similar predicates to entities which themselves again refer to identical (or similar) places, events, actors, or objects [11]. In case of York, two sources which describe the occupation by the Danish army, the arrival of a person referred to by *William the Conqueror*, and so forth, but use different place names (e.g., Yórvik and Yorke) probably still refer to the same city of York. This approach is not restricted to large-scale places such as cities but also works on the level of single streets and properties. Using semantic similarity reasoning for identity assumptions is drawing an analogy from our geographical notion of place to the place within

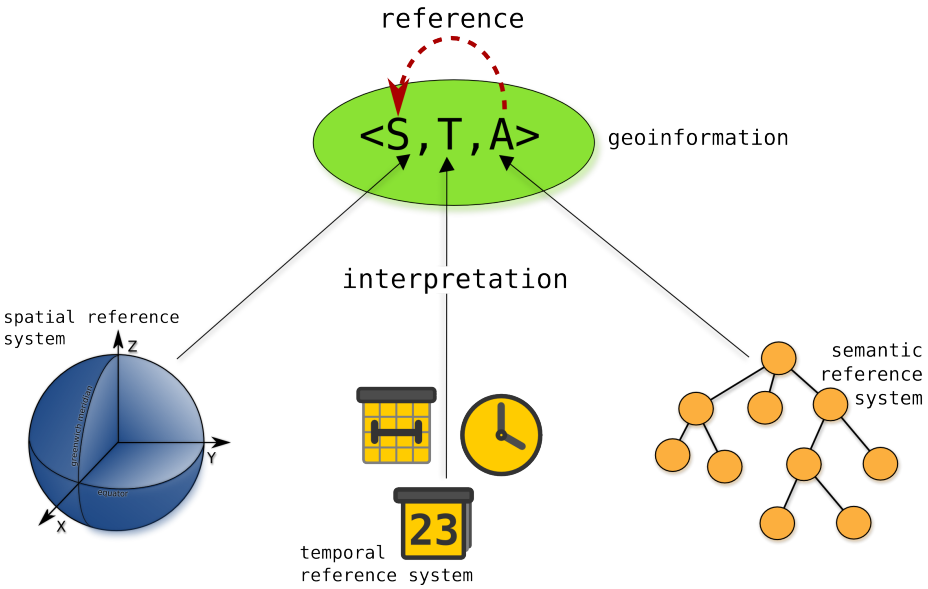

Figure 3: The interpretation of the **S**patial, **T**emporal, and **A**ttributive/thematic data using reference systems. Thematic data can be used as additional source of reference if spatial (or temporal) data is vague or insufficient [11].

a network of historical facts and the spatial *next-to* relation to a thematic one based on similarity estimations [11, 13].

3 Structuring Microtheories using Space and Time

While the first section discussed different perspectives on how to model places in a heterogeneous setting such as the envisioned ppGIS, the second section proposed to use thematic data as additional source of reference if classical spatial reference is missing, vague, or misleading. Consequently, while we focused on particular types and individuals so far, the last section focuses on conceptual reference models such as ontologies.

As argued above, CIDOC CRM intentionally does not implement rigid constraints for the use of types and relations. Such an approach has many benefits for annotating vague and incomplete historical data but at the same time restricts automatic-reasoning and complex queries. For instance, we cannot infer a contradiction from the fact that a person participated in two spatially disjoint events at the same time. The envisioned ppGIS should be able to organize and provide access to data in different formats, from different ages, but especially also from different contributors. Some of these sources may require different ontological assumptions and conceptualization than others. For instance, the system should be able to store information about actors even if their (real) existence is questionable or their live spans are of an unrealistic duration. At the same time, the ppGIS should be used by scholars to discover and infer new facts and to clean up external knowledge bases. Finally, opening the system for

volunteered geographic information may introduce other requirements [16].

A promising approach to deal with such a heterogeneous setting are microtheories (also called contexts) as implemented by the (Open)CyC ontology. Microtheories follow the classical idea from artificial intelligence research to be locally consistent but allow for inconsistencies in the total (global) knowledge base. With respect to the ppGIS, this allows to handle different (and even contradictory) conceptualizations used to describe historical facts. One microtheory may be strict about the nature of human beings (e.g., forcing the user to link each person to exactly one birth events), while other local microtheories may have weaker constraints. Usually microtheories are organized by subsumption hierarchies. All facts in the super-theory must hold for each of its sub-theories while sibling-theories may contain different conceptualizations. A promising direction for further research may be to introduce time and space as additional first class ordering principles for microtheories (see Figure 4). For instance, theories developed to store and reason about historical documents from the Middle Ages may be organized in another branch of the knowledge base than microtheories dealing with data from the age of industrialization. Similarly to the temporal example, space can also be used to structure microtheories. The definition of rivers, for instance, differs clearly between southern European and northern European countries. In contrast to the classical subsumption/generalization case depicted in Figure 4a, 4b and 4c combine temporal and spatial inclusion with subsumption.

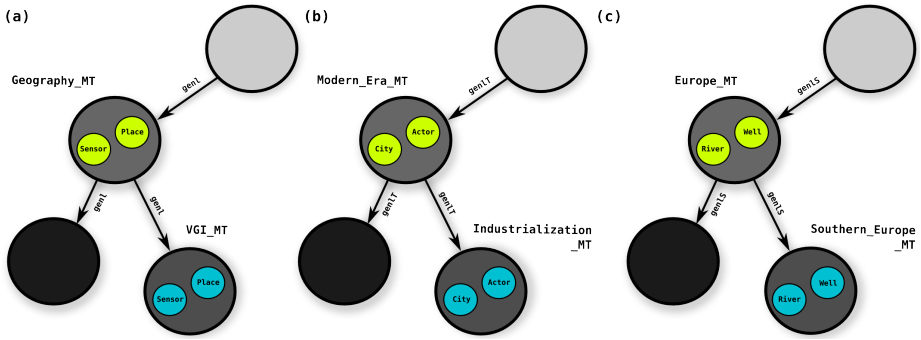

Figure 4: Structuring microtheories by (a) generalization (*genl*), (b) generalization and temporal inclusion (*genlT*), and (c) generalization and spatial inclusion (*genlS*).

In this context, one important question is how to define the level of granularity for these microtheories. In general, the author believes that the role of ontology is splitting up the world in parts, while reasoning is responsible for putting it together again (and make implicit information explicit).

4 Conclusions

In this paper we have discussed several challenges for the representation of and reasoning about geographic places in heterogeneous settings. We have argued that places are social constructs and could be modeled using different paradigms. Place names, types, and spatial footprints which act as the pillars of gazetteers

are just three prominent characteristics of places often used for referencing. Non of them needs to persist through the existence of a place nor are they unique in their referencing ability. In their absence thematic knowledge can take over this role. Instead of geographical location interpreted through spatial reference systems, the location within a network of historical facts is used to reference and disambiguate places based on conceptual reference systems such as CIDOC CRM. Historical facts, however, have a temporal scope and may be even contradictory. Similarly, the definitions of classes and relations change over time. To reflect this, the ontologies underlying the public participatory GIS can be structured as microtheories. This allows to be consistent on the local level, e.g., within particular epochs, without the need to enforce global definitions.

Finally, if the cultural heritage data stored by the ppGIS would be available as Linked Spatiotemporal Data (as RDF triples) together with the ontologies it could be directly embedded into third party applications and webpages to generate context and user-aware information on the fly using technologies such as *JSExplicit*[3]. Similarly, one could use the data to develop a linked data gazetteer (see [8, 16] for details).

Acknowledgments

This research has been funded by the International Research Training Group on *Semantic Integration of Geospatial Information* (DFG GRK 1498) and the *Simcat II* project (DFG JA1709/2-2). The author would also like to thank the organizers of *The Cultural Heritage of Historic European Cities and Public Participatory GIS* workshop for their invitation.

References

[1] Barsalou, L., Sloman, S., and Chaigneau, S. *Representing functional features for language and space: Insights from perception, categorization and development.* Oxford University Press: New York, 2005, ch. The HIPE Theory of Function, pp. 131–147.

[2] Bennett, B., and Agarwal, P. Semantic categories underlying the meaning of 'place'. In *Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings* (2007), vol. 4736 of *Lecture Notes in Computer Science*, Springer, pp. 78–95.

[3] Casati, R., and Varzi, A. C. *Parts and Places. The Structures of Spatial Representation.* MIT Press, Cambridge and London, 1999.

[4] Chrisman, N. *Exploring Geographical Information Systems.* Wiley, 2001.

[5] Cresswell, T. *Place: A Short Introduction.* London: Blackwel, 2004.

[6] Crofts, N., Doerr, M., Gill, T., Stead, S., and Stiff, M. *Definition of the CIDOC Conceptual Reference Model (Manual Version 4.2.1)*, October 2006.

[3] See jsexplicit.sourceforge.net for details.

[7] Goodchild, M. Geographical data modeling. *Computers and Geosciences* (1992), 401–408.

[8] Hart, G. Linking to the past, geographically speaking: The linked data web and historical gis. In *The Cultural Heritage of Historic European Cities and Public Participatory GIS Workshop. 17–18 September 2009* (2009).

[9] Hill, L. L. *Georeferencing: The Geographic Associations of Information (Digital Libraries and Electronic Publishing)*. The MIT Press, 2006.

[10] Hill, L. L., Frew, J., and Qi, Z. Geographic names: The implementation of a gazetteer in a georeferenced digital library. *D-Lib Magazine Vol. 5, No. 1* (1999).

[11] Janowicz, K. Towards a similarity-based identity assumption service for historical places. In *Geographic Information Science, 4th International Conference (GIScience 2006)* (Münster, Germany, September 2006), M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild, Eds., vol. 4197 of *Lecture Notes in Computer Science*, pp. 199–216.

[12] Janowicz, K., and Kessler, C. The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science 10*, 22 (2008), 1129–1157.

[13] Janowicz, K., Kessler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., and Baeumer, B. Algorithm, implementation and application of the SIM-DL similarity server. In *Second International Conference on GeoSpatial Semantics (GeoS 2007)* (Mexico City, Mexico, November 2007), F. T. Fonseca, A. Rodríguez, and S. Levashkin, Eds., no. 4853 in Lecture Notes in Computer Science, Springer, pp. 128–145.

[14] Jordan, T., Raubal, M., Gartrell, B., and Egenhofer, M. J. An affordance-based model of place in gis. T. Poiker and N. Chrisman, Eds., IUG, pp. 98–109.

[15] Kauppinen, T., Henriksson, R., Sinkkilä, R., Lindroos, R., Väätäinen, J., and Hyvönen, E. Ontology-based disambiguation of spatiotemporal locations. In *Proceedings of the 1st international workshop on Identity and Reference on the Semantic Web (IRSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008)* (2008), P. Bouquet, H. Halpin, H. Stoermer, and G. Tummarello, Eds., vol. 422 of *CEUR Workshop Proceedings*, CEUR-WS.org.

[16] Kessler, C., Janowicz, K., and Bishr, M. An agenda for the next generation gazetteer: Geographic information contribution & retrieval. In *International Conference on Advances in Geographic Information Systems 2009 (ACM SIGSPATIAL GIS 2009)* (2009; forthcoming).

[17] Kuhn, W. Semantic engineering. In *Research Trends in Geographic Information Science* (2009; in press), G. Navratil, Ed., Springer.

[18] Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. WonderWeb Deliverable D18 –Ontology Library (final). Tech. rep., Laboratory For Applied Ontology - ISTC-CNR, 2003.

[19] Sowa, J. F. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA, USA, 2000.