Building Place Ontologies for the Semantic Web: Issues and Approaches

A.I. Abdelmoty
School of Computer Science,
Cardiff University,
Wales, UK
a.i.abdelmoty@cs.cf.ac.uk

P. Smart
School of Computer Science,
Cardiff University,
Wales, UK
p.smart@cs.cf.ac.uk

C.B. Jones
School of Computer Science,
Cardiff University,
Wales, UK
c.b.jones@cs.cf.ac.uk

ABSTRACT

Place geo-ontologies have a key role to play in the development of the geospatial-semantic web, with regard to facilitating the search for geographical information and resources. They normally hold large amounts of geographic information and undergo a continuous process of revision and update. This papers reviews the limitations of the OWL ontology language for the representation of Place and proposes two novel approaches to frameworks that combine rules and OWL for building and managing Place ontologies.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms: Design,Algorithms,Languages

Keywords: Geo-IR, Semantic Web, Rules, OWL, Geographic Ontologies,Mediation

1. INTRODUCTION

Interest in geographic information, and in particular information related to Places and Place names, has grown significantly over the past few years. The powerful simplicity of applications such as Google Earth fueled a wealth of geo-related activities and needs for web users. Also, many initiatives are ongoing to build national and global spatial data infrastructures to enable the share, use and reuse of geographic information. Searching and retrieval of Place-related information is central to these activities. An essential component of search engines that support the effective retrieval of geographically referenced resources are Place ontologies. These are models of terminology and structure of geographic space as well as records of entities in this space [12].

Building Place ontologies is a non-trivial task that involves the integrated utilisation of heterogenous spatial data resources. The complexity of the problem is related to the inherent complexity and dimensions of the data. A Place may have more than one name, may be related to multiple place concepts and types, as well as associated with multiple spatial representations of its location and shape. Much of the semantics in Place ontologies are implicit and evident only at the instance level. For example, different types of spatial relationships exist between Places; a Place may be inside, north-of, near to, larger than another, etc. Some of these relationships may be captured on the concept level but most others are implicit, evident only by visual interpretation and geometric computation. Explicit representation of such relationships is not practically possible and means for their automatic extraction are needed.

Maintaining the logical as well as spatial integrity of Place ontologies is crucial for maintaining their soundness and viability. Spatial integrity is different from logical integrity and is not directly implied by it. For example, a part-of semantic relationship between two geo-objects does not imply directly the correct relationships between the objects' spatial representations. The boundary of the child object might intersect with the parent, or the area of the child might be larger than the parent, etc.

This paper starts by identifying the limitations of the OWL ontology language for representing Place ontologies and then proposes two frameworks for the development of Place ontology management systems. The first approach assumes a centralised view of ontology development, where the instance store (or ABox - assertion box which records observations of the world [3]) is populated from available data sources. In the second approach, no (or limited) instance store is assumed and the Place information is derived from the integration of multiple data resources. Both frameworks employ a combination of qualitative and quantitative spatial reasoning over the Place ontologies and hence assumes the need for combining OWL with rules. An overview of both approaches is given. A detailed development of a spatial rule-markup language to support combining OWL and rules is given in [19].

1.1 Rules for Place ontologies

Rule expression over Place ontologies is needed for the representation of the following types of rules:

- Spatial reasoning rules for the deduction of implicit geo-semantics.
- Spatial integrity rules for representing different type of spatial integrity constraints to maintain the consistency of Place ontologies.

In [2] an evaluation of OWL as a language for representing geo-ontologies identified several limitations, including,

GIR'07, November 9, 2007, Lisboa, Portugal.

no support for the computation of properties and comparison of individuals. OWL's flat file XML representation can be inefficient when dealing with large geometric data sets and in general, OWL and RDF(S) do not support all the necessary semantics for processing geo-spatial data. In addition, the following limitation are noted.

1. OWL's first order, open world semantics in combination with the non unique name assumption is not suitable for constraint checking [5]. Extensions to OWL have been proposed to overcome this limitation, for example by translating subsets of OWL to a logic program that assumes both unique name and closed world assumptions[5, 8].

2. OWL can't be used to represent inference patterns of the form, $\forall x, y, c : rel_1(x, y) \wedge rel_2(y, c) \rightarrow rel_3(x, c)$, so called triangular knowledge [10]. This is a typical form of a spatial reasoning rule for composition of spatial relationships.

3. OWL does not support spatial data types. This leads to a poor representation of geometric objects using generic class and property constructs with potentially high storage overheads [9].

4. OWL does not support geometric processing, computation or spatial indexing and hence it is difficult to perform simple computations over geometries, such as, area or distance.

5. Tableaux based reasoners are poor for query answering over individuals [5]. Instance bases of Place ontologies are likely to be very large, hence, logic programming reasoning engines are more appropriate in this case.

Efforts are ongoing to store OWL-based ontologies in traditional relational databases such as PostgreSQL [1]. The techniques involve the representation of the RDF triple data model using the database structure. Reasoning over the ontologies is however, not supported in conventional databases, and instead the ontologies are loaded into and reasoned with in memory by an appropriate inference engine. This solution is not practical with large Place geo-ontologies where substantial amounts of memory would normally be required.

Most of the storage overhead is related to the coordinate location information associated with the features in the ontology. During real world experimentation with the SABE [16] data-set, a total of 10959 individuals (only a small subset of the total number of individuals that SABE contains) where converted to an OWL geo-ontology. Without locational information the geo-ontology consumed 2.2 mb of persistent storage and ,when reasoned with, 16 mb of main memory. Adding locational information expanded the ontology to consume 100 mb of persistent storage and over 1 gb of main memory when reasoned with . Also, as noted above, the type of geometrical computation operations, such as distance or area, required to operate over locational information are not supported using OWL schema or functions. Hence, it is more appropriate to delegate the representation and management of the absolute locational information to an external geometric processor or a spatial database system.

Recently, rule languages have been proposed that complement and enhance the expressiveness of standard ontology language. SWRL [11] is an extension of the OWL semantics with a subset of RuleML that allows augmenting OWL ontologies with horn clauses and thus horn logic. RuleML [15] provides a standardized vocabulary for the specification of rules and has now established itself as the *interlingua franca* of web based rule languages. An approach to the use of SWRL for maintaining the integrity of spatial data acquired by mobile devices is proposed in [14]. Laser-scan's Radius Studio [2] uses SWRL based rules to define and execute business rules over spatial data.

Rule languages require a corresponding rule engine in order to operate. One of the most complete toolkits that incorporates creating and manipulating ontologies with a powerful reasoning engine is Jena2 [13]. Other rule systems that incorporate both rule language and engine have been proposed, for example, Algernon [4], Euler [3] and JDrew [4].

Geo-logica [20] SRI International's framework for query answering across multiple external resources, is a promising approach to geographic multi resource mediation.The framework incorporates both spatial and temporal reasoning for query answering using the SNARK deductive first order logic theorem prover. SNARK composes multiple sources through agents (procedural attachments) that are accessed on the fly during query answering. The Geo-logic framework is principally suited for deductive reasoning incorporating disjoint resource types, and does not consider the integration, selection and maintenance of overlapping resources.

2. A CENTRALISED APPROACH TO PLACE ONTOLOGY DEVELOPMENT

In this section, a new framework is proposed for the representation, storage and management of a Place geo-ontology. The framework assumes a centralised view where the Place ontology evolves by integrating new data to a central store of terminology and instances. A hybrid approach to the representation of Place information is adopted, where OWL is used to represent the concept structure and hierarchies of Place and an external geometric data store handles the spatial representation of place footprints. In addition, a spatial reasoning engine is incorporated that allows for the expression of rules and their implementation over the Place ontology. This approach has been successfully prototyped . The architecture is shown in Figure 1 and its components are described below.

2.1 Place Ontology Management System

The Place ontology subsystem combines both the OWL Place ontology management system and the geometric footprint data store and management system.

The OWL Place Ontology subsystem.

The basic concept in a Place ontology is a geographic Place that is normally associated with a name (toponym), one or more alternative names and one or more geometric representations of its location (footprint). Semantic associations as well as spatial relationships can be modeled between

---

[1] http://www.postgresql.org/

[2] http://www.laser-scan.com/technologies/enterprise/radius studio/index.htm2004

[3] http://www.agfa.com/w3c/euler/

[4] http://www.jdrew.org/jDREWebsite/jDREW.html

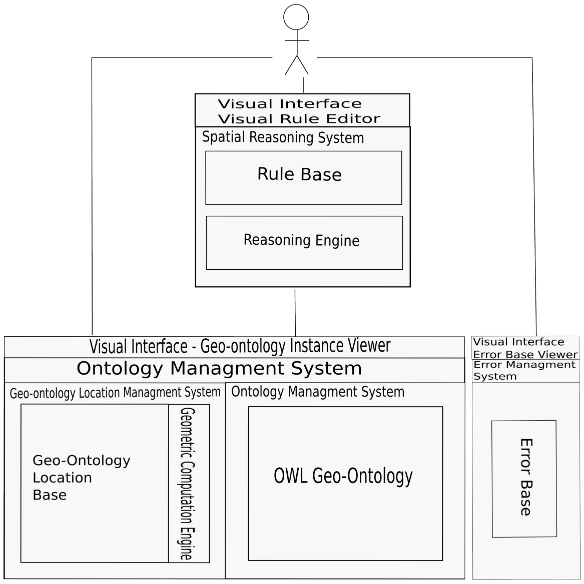


Figure 1: A centralised approach to Place Ontology Development

place concepts. The Open GeoSpatial Consortium (OGC) [1] standard spatial data model specification is assumed for representing the spatial properties of Place and its relationships. A Place ontology model is shown in Figure 2.

Place Footprint subsystem.

The Place footprint subsystem is a GIS or a spatial database system used for the storage and manipulation of the geometric location(s) of a Place. This system exposes powerful features of geometric data processing and spatial indexing, that can't be readily implemented over OWL ontologies, to the framework. A unique reference between features in the OWL ontology and their corresponding footprint(s) information is maintained. URIs are used to provide a unique identifier (primary key) in the footprint system and link to the features of the OWL ontology. The footprint system only interacts and is accessible through the the main ontology system and one interface is used in the framework for both subsystems.

The following is an example of the OWL syntax used to represent a Place and a spatial relationship from the model in Figure 2.

```
<Region rdf:ID="http://geo.ont/1232Wales">
      <Description rdf:datatype=
              "http://www.w3.org/2001/XMLSchema#string"
               >A region</Description>
      <StandardToponym>
           <Toponym rdf:ID="Toponym_1">
           <Name rdf:datatype=
                 "http://www.w3.org/2001/XMLSchema#string"
                 >Wales</Name >
           <Language rdf:datatype=
                 "http://www.w3.org/2001/XMLSchema#string"
                  >English</Language>
       </StandardToponym >
       <PlaceType>
           <Geo-PlaceType rdf:ID="CountryType">
           <FeatureTypeName rdf:datatype=
                  "http://www.w3.org/2001/XMLSchema#string"
                   >Country</FeatureTypeName >
       </PlaceType >
      <Adjacent>
        <Region rdf:ID="http://geo.ont/134England">
                ...
        </Region>
      </Adjacent>
      <Language rdf:datatype=
            "http://www.w3.org/2001/XMLSchema#string"
             >English</Language>
      <Date rdf:datatype=
             "http://www.w3.org/2001/XMLSchema#date">
              2006-09-30</Date>
</Region>
```

The geometry of the Place is stored in a spatial database system for example Oracle Spatial 10g. Oracle has a set of spatial schemas for the definition and representation of spatial objects. In Oracle, a table is defined for a Place object whose primary key is the URI or the RDFID. The geometric description of the object is then stored using the system's specific representation as MDSYS.SDO_GEOMETRY. An MDSYS_SDO_GEOMETRY object is part of the Oracle spatial schema used for the representation of different geometric data types.

Calls to geometric data processing functions in Oracle are made, as needed, using the rule engine with built-in spatial predicates that assumes the unique association with the OWL place individuals.

2.2 Spatial Reasoning System

Rules over geo-ontologies are used to allow for the automatic derivation of implicit spatial information and for expressing spatial integrity constraints to maintain the spatial consistency of the ontology.

A possible classification of the types of spatial rules to be represented by the engine is as follows.

- Rules representing constraints over object properties in space, in particular, spatial properties of dimension, shape and size. Examples of these types of rules include the fact that a polygon must have at least three different points and that a polygon must be closed, etc. These types of constraints are normally used in spatial databases and GIS.

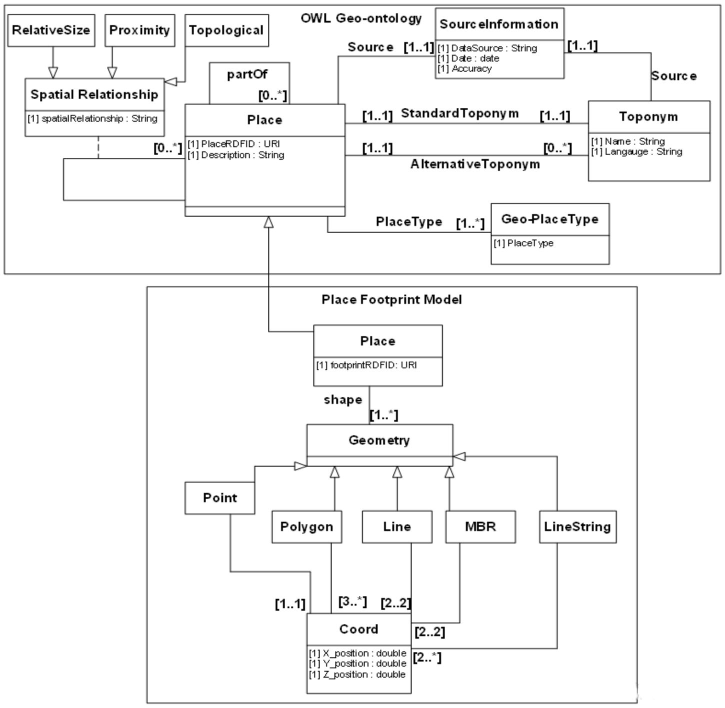


Figure 2: OWL Geo-ontology and GIS Location Base Model

- Rules for reasoning over spatial relationships between objects in space. For example, the fact that an object A is located inside another object B and that B is inside object C, implies that object A is also inside C. It also implies that C is larger than A and B. This is an example of qualitative spatial reasoning (QSR). The spatial reasoning engine utilise the results of the large body of research in this field, where automated methods have been proposed for the derivation of spatial composition tables for different types of spatial objects and relationships.

A spatial reasoning engine is developed as an extension to Jena. Jena is an open source Java-based semantic web tool that can be used to reason over OWL using OWL's standard inference mechanisms as well as allows for the expression and implementation of user-defined rules over OWL. Jena implements the rules using a RETE-based forward production rule system [18, 6], and an XSB [17] based backward chaining logic programming engine. The spatial reasoning engine proposed here extends the standard Jena architecture by the following features.

1. Interleaved Rule Execution. Forward and backward reasoning modes can be intermixed - antecedents of a forward rule can be queried using a backward rule during runtime. This feature is required to allow for on-the-fly reasoning over spatial relationships. Spatial relationships may need to be computed using the geometric data store if they are not stored explicitly in the ontology.
2. A restricted form of Courteous Logic [7] is used to provide support for the expression of rule exceptions and rule priorities to indicate the order of rule execution in the case of conflicts.
3. Spatial built-ins are defined and computed in the geometric data store. Examples include area, distance, containment and adjacency relationships.
4. Spatial rule meta-data and a spatial syntax for help in the visual management and authoring of rule bases.

The framework is capable of expressing spatial reasoning rules as shown in the following example.

$$\begin{aligned}[&Region(?x) \wedge Region(?y) \wedge Region(?c) \quad \wedge \\ &\qquad Inside(?x?c) \wedge Inside(?c?y) \quad \rightarrow \quad Inside(?x?y)]\end{aligned}$$

The conclusion `Inside(?x ?c)` would only be inferred if both the atoms `Inside(?x ?c)` and `Inside(?c ?y)` can be satisfied. These atoms are either satisfied by facts directly stored in the ontology (explicit), or inferred using spatial relationship composition rules (implicit), or as a last resort satisfied by a rule that calls an external geo-computation engine.

For example, the following is a subset of rules used to derive the inside relationship between two regions. The last rule is a call to external geo-computation (*exInside* predicate).

$$\begin{array}{lcl}
Inside(?x\ ?y) & \leftarrow & Region(?x) \wedge Region(?y) \wedge Region(?c) \\
 & \wedge & Inside(?x\ ?c) \wedge Equal(?c\ ?y) \\
Inside(?x\ ?y) & \leftarrow & Region(?x) \wedge Region(?y) \wedge Region(?c) \\
 & \wedge & Inside(?x\ ?c) \wedge Inside(?c\ ?y) \\
Inside(?x\ ?y) & \leftarrow & Region(?x) \wedge Region(?y) \wedge Region(?c) \\
 & \wedge & Inside(?x\ ?c) \wedge CoveredBy(?c\ ?y) \\
Inside(?x\ ?y) & \leftarrow & Region(?x) \wedge Region(?y) \wedge Region(?c) \\
 & \wedge & CoveredBy(?x\ ?c) \wedge Inside(?c\ ?y) \\
Inside(?x\ ?y) & \leftarrow & Region(?x) \wedge Region(?y) \wedge Region(?c) \\
 & \wedge & exInside(?c\ ?y)
\end{array}$$

By interleaving forward and backward reasoning modes, facts can be derived (using qualitative reasoning), or proven, on the fly by a set of one or more backward rules (quantitative reasoning or geometric computation) as shown in the above example. The obvious benefits from the combined mode is the reduction of storage and computational overheads.

The framework also allows for the definition of user-defined rules and rule exceptions, as shown in the following example.

$$\begin{aligned}
&Road(?x) \wedge River(?y) \wedge Crosses(?x?y) \rightarrow \\
&error(roadRiverCrossError?x\ Crosses\ ?y \\
&roadsRiversNotCross\ riverRCross) \qquad (1) \\
&Road(A40) \wedge River(Taff) \wedge Crosses(A40\ Taff) \rightarrow \\
&notError(roadRiverCrossError\ A40\ Crosses\ Taff \\
&roadsRiversCross\ riverRCrossException) \qquad (2)
\end{aligned}$$

Rule (1) is the default rule and (2) is its exception. Intuitively, the ground instantiation of the first rule which substitutes variables ?x and ?y for A40 and Taff, respectively is overridden by the second rule.

2.3 Error Management System

Errors detected by the spatial integrity rules are stored in a separate error ontology. A trace of the type of errors as well as the spatial integrity rules fired to detect the error is maintained. Analysis of the error ontology can give some insight to types of integrity problems in the data and their frequency. This can be useful in guiding the development of the error management system.

3. DISTRIBUTED APPROACH TO PLACE ONTOLOGY DEVELOPMENT

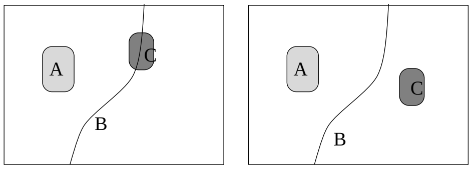
In contrast to the centralised approach, the Place ontology here, which is currently under development, assumes access to external data sources for Place information. I.e. the ontology will possibly not be populated directly from these sources, but can query and access the information as required. It is likely that disparate sources hold different types of Place information in varying qualities and granularities. Some sources may be freely available on the Web, for example, geonames.org, while others are offered on pay-per-access schemes, etc.

In this framework, the Place ontology system acts as a mediator between users of the Place information and resources of Place information. The ontology system maintains information on the data sources they include, their geographic coverage, their completeness, with respect the areas covered, the types of geographic places, and the granularity and accuracy of representation. Some or all of this information may be available directly as metadata for the sources or may need to be extracted by the ontology system.

A specification of a request from users of the Place ontology system invokes a search request by the system to the data sources available. A ranked list of possible Place objects (or references to them) is then returned to the user. The ranking can be based on the criteria specified by the user, e.g. level of detail, spatial representation, types of attributes as well as on the qualities of the data sources, e.g. availability and currency of the information.

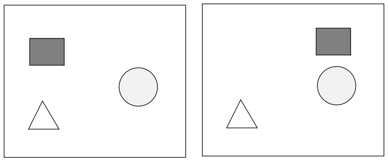
The framework assumes a service-oriented approach where the data sources are invoked through requests to web service interfaces. Central to the framework is a resource management system that transfers requests to and interprets responses from resources. The architecture of the system is shown in Figure 3.

Figure 4: Topological inconsistency. (a) Object B crosses object C. (b) Object B is disjoint from C.



(a) (b)

Figure 5: Directional inconsistency.



(a) (b)

As information is requested through the resource manager, it is indexed both spatially and semantically. The evolving spatial index will track which resources contain place information in a certain location in space and the semantic index traces the types of place information these resources hold. The ontology instance store will, in general, refer to place identifiers and their resources and not actual places and their footprints. Although in some cases, caching of frequently requested place information would be appropriate.

The spatial reasoning and geometric computation subsystem plays an important role in maintaining the ontology as well as the place indexes built. In addition, it is used by the resource management component to determine the interpret the quality of the Place information extracted from the different resources.

The problem of combining various place name resources involves filtering and integration at both the semantic and the spatial levels. It is necessary to establish equivalences between instances of named places in different data sets. The same place may be allocated different classifications by different organisations and thus interpreting and mapping across place type hierarchies is required. Attempts to determine the equivalence of the geometric location can be hindered by the wide variation in the accuracy and precision of the coordinates as well as variation in the types of coordinate system. Hence, different types of absolute and relative spatial similarity measures need to be devised. Figure 4 and Figure 5 show examples of possible spatial similarity tests.

4. CONCLUSIONS

Limitations of the OWL ontology language have been identified in the literature and rule languages are currently being developed to complement ontology representation for the semantic web. In this paper, we discussed the particular requirements of Place geo-ontologies and the need for combining OWL and spatial reasoning rules to support their development and maintenance. Two frameworks for building Place ontologies are proposed. The first approach assumes a centralised view where a place ontology instance store evolves by importing data from external resources. In

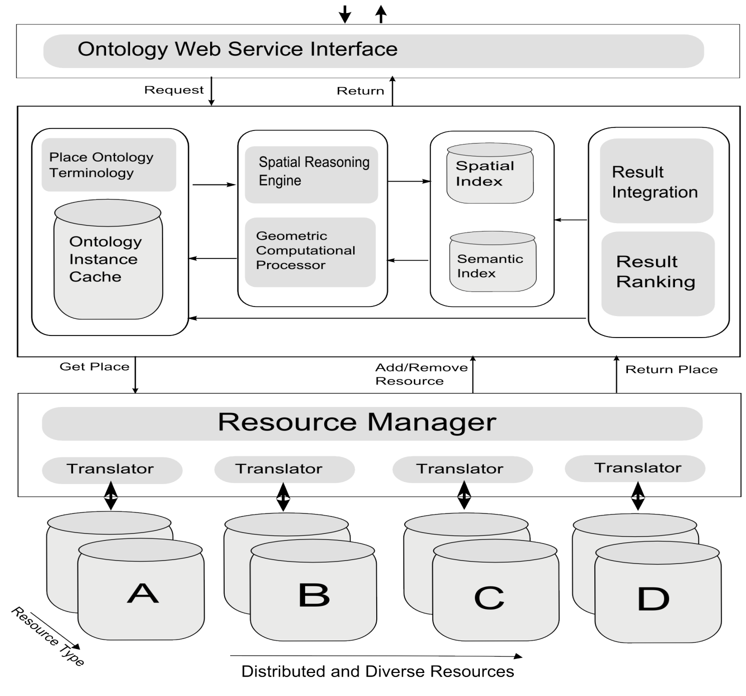


Figure 3: Distributed Place Geo-ontology framework

the second approach, the ontology instances are pointers to references of Place information in diverse data resources. The first approach is more suited to scenarios where strong control and ownership of the data is envisaged, while the second approach is more suitable in more general situations of information retrieval from data sources of diverse quality and granularity of content.

5. ACKNOWLEDGMENTS

The research is supported by the European commission FP6 project TRIPOD: TRI-Partite multimedia Object Description and its funding is duly acknowledged.

6. REFERENCES

[1] Open gis consortium topic 5 features. Internet, 2005. http://portal.opengeospatial.org/files/?artifact_id=890.

[2] Abdelmoty, A. I., Smart, P. D., Jones, C. B., Fu, G., and Finch, D. *A critical evaluation of ontology languages for geographic information retrieval on the Internet.* Journal of Visual Languages & Computing. August 2005, pp. 331–358.

[3] Belleghem, K. V., Denecker, M., and Schreye, D. D. A strong correspondence between description logics and open logic programming. In *International Conference on Logic Programming* (1997), pp. 346–360.

[4] Crawford, J. M., and Kuipers, B. J. Algernon a tractable system for knowledge-representation. *SIGART Bull. 2*, 3 (1991), 35–44.

[5] de Bruijn, J., Lara, R., Polleres, A., and Fensel, D. Owl dl vs. owl flight: conceptual modeling and reasoning for the semantic web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web* (New York, NY, USA, 2005), ACM Press, pp. 623–632.

[6] Forgy, C. L. Ops5 reference manual. Tech. Rep. CMU-CS-81-135, Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA, 1981.

[7] Grosof, B. N. Prioritized conflict handling for logic programs. In *ILPS* (1997), pp. 197–211.

[8] Grosof, B. N., Horrocks, I., Volz, R., and Decker, S. Description logic programs: combining logic programs with description logic. In *WWW* (2003), pp. 48–57.

[9] Haarslev, V., Lutz, C., and Moller, R. Foundations of spatioterminological reasoning with description logics. In *Principles of Knowledge Representation and Reasoning* (1998), pp. 112–123.

[10] Horrocks, I. Owl rules, ok? In *Rule Languages for Interoperability* (2005).

[11] Horrocks, I., Patel-Schneider, P. F., Tabet, H. B. S., Grosof, B., and Dean, M. Swrl: A semantic web rule language combining owl and ruleml. Internet Report, May 2004. http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/.

[12] Jones, C., Abdelmoty, A., and Fu, G. Maintaining ontologies for geographical information retrieval on the web, 2003.

[13] McBride, B. Jena: A semantic web toolkit. Internet, November 2002. http://dsonline.computer.org/0211/f/wp6jena.htm.

[14] Ms, Stephan; Wang, F. R. W. Using ontologies for integrity constraint definition. In *4th International Symposium On Spatial Data Quality* (2005), no. ISBN 962-367-473-2.

[15] RuleML. Ruleml web site. Internet, 2006. http://www.ruleml.org/.

[16] SABE. Sabe: Seamless administrative boundaries of europe. http://www.ordsvy.gov.uk/products/computer/sabe/index.htm.

[17] Sagonas, K., Swift, T., and Warren, D. S. Xsb: An overview of its use and implementation. Tech. rep., Nov. 02 1993.

[18] Schneier, B. The rete matching algorithm. *AI Expert 7*, 12 (Dec. 1992), 24–29.

[19] Smart, P., Abdelmoty, A., El-Geresy, B., and Jones, C. A Framework for combining Rules and Geo-ontologies. In *Proceeding of the First International Conference on Web Reasoning and Rule Systems* (2007), vol. LNCS 4524, Springer, pp. 133–147.

[20] Waldinger, R., Appelt, D. E., Fry, J., Israel, D. J., Jarvis, P., Martin, D., Riehemann, S., Stickel, M. E., Tyson, M., Hobbs, J., and Dungan, J. L. *Deductive Question Answering from Multiple Resources.* AAAI, Menlo Park, CA, 2004.