Concept Level Matching of Geospatial Ontologies

A methodology and tool support for semi-automatic alignment of heterogeneous geospatial ontologies

Isabel F. CRUZ, William G. SUNNA and Kalyan AYLOO

The problem of querying geospatial databases in a distributed environment can be resolved by aligning the ontologies that describe each database. The alignment process consists of establishing mappings between related concepts in the ontologies. Once such mappings are established, data can be integrated. In particular, a query written in terms of an ontology can be translated into queries on the other ontologies. We discuss alignment in two different architectures—centralized and peer-to-peer—and give real-world examples of geospatial data to illustrate our alignment process. The alignment process is based on certain mapping types. Our alignment methods are semi-automatic and are supported by a tool. We give experimental results in which we show that approximately one fifth of the mappings can be established automatically using our method.

KEYWORDS

Ontology-based data integration, Ontology alignment, Centralized data system, Peer-to-peer data system, Geospatial data, Land use data, Wetland classification.

INTRODUCTION

In this paper we describe an ontology-based approach to bridge the semantic gap among distributed geospatial data. We assume that there is an ontology associated with each database that describes the hierarchical relationship between the database concepts. Because such ontologies are created independently from one another, they are heterogeneous. In particular, we focus on semantic heterogeneity, which arises when different terms are used to describe the same concept or when the same term is used to describe different concepts [3]. Our contribution is to the alignment process, which enables experts to establish mappings between related concepts in different ontologies, thus allowing for the integration of data across multiple databases. For example, using those mappings, a query to one of the databases can be translated to a query to any of the other databases. We present two architectures: (i) a centralized integrated system, and (ii) a peer-to-peer integrated system.
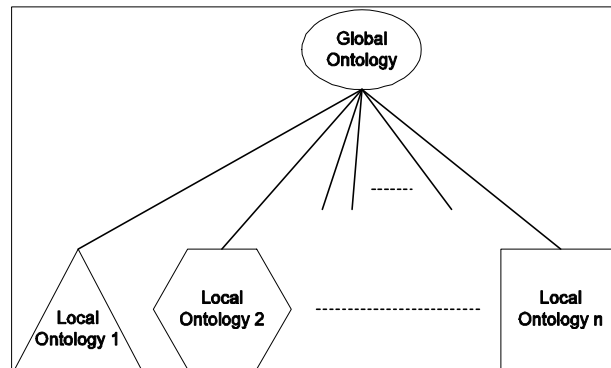

Figure 1: Centralized integrated system.

In a centralized integrated system, a global ontology is introduced as shown in Figure 1. Each distributed database that wishes to participate in the integrated system must have its ontology

aligned with the global ontology. We refer to the ontologies of the distributed databases as *distributed* or *local* ontologies, to distinguish them from the global ontology. Several cases can occur in the creation of a global ontology. For example, the global ontology can be designed so as to encompass as much as possible the information contained in the distributed ontologies. Another possibility is that a group of domain experts assembles an ontology for a particular domain, which could become a standard for that domain. In this paper, we consider the case where the global ontology is created independently from the other ontologies, thus corresponding to the "local as view approach" in data integration [4]. In this way, the addition (or removal) of a database will not affect the global ontology.

In a centralized integrated system, a query to the distributed databases is posed in terms of the global ontology. The alignment process establishes mappings between the global ontology and each distributed ontology, thus allowing for the translation of a query on the global ontology into a query on a distributed ontology.

To demonstrate how ontology alignment is used in a centralized architecture, we provide a data integration example for land use in the state of Wisconsin. In this case, the distributed ontologies were separately developed and are quite dissimilar.
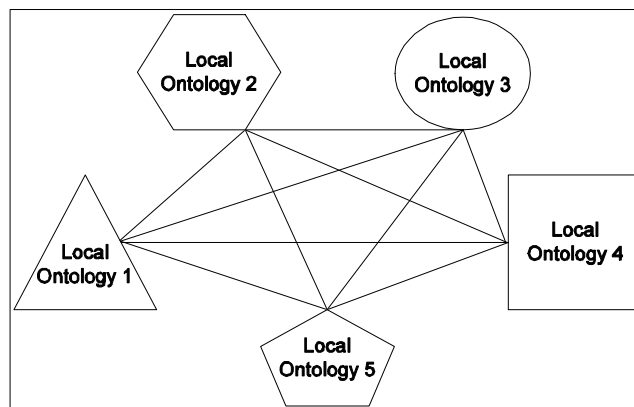

Figure 2: Five databases participating in a peer-to-peer integrated system.

In a peer-to-peer integrated system, a query posed on one of the databases (the *target database*) is propagated to the others in the network as depicted in Figure 2. The query is posed in terms of the ontology of the target database and will be translated to a query expressed in terms of another database, using the mapping of concepts that was established through the alignment process.

For a peer-to-peer architecture, we concentrate on wetland classifications and in particular on an established standard, the "Cowardin" Wetland Classification [5]. We will use this standard together with its variant, the South African Wetland Classification Inventory [7] and establish their alignment. In a peer-to-peer framework, queries on the former will be translated to queries on the latter, using the mappings that were established in the alignment process. This example also demonstrates that even when standards are available, it is likely that no two exact copies exist. Any variation that is introduced will not permit the simple integration of the data, in that a query on one of the databases cannot be applied, without modification, to the other database.

To assist users in the alignment process, we implemented a tool whose output is a set of mappings between the aligned ontologies. Our tool supports semi-automatic methods that can deduce mappings between concepts thus helping users make accurate decisions and speeding up the mapping process.

The paper is organized as follows. In the section on *Related Work*, we give an overview of previous

and current research in the area. We present land use codes and wetland classification ontologies in the *Geospatial Ontologies* section. In the *Ontology Alignment* section, we present the type of mappings we consider in performing the alignment between two ontologies and describe our approach to automatically propagate mappings along the two ontologies involved in the alignment process. We describe the implementation of our alignment tool that supports the semi-automatic alignment process in the section on the *Agreement Maker*. In the *Experimental Results* section, we describe an experiment we conducted with our alignment tool based on a real world example from the wetland classifications domain. Finally, in the *Conclusions* section we propose ways of extending our alignment process and outline future work.

RELATED WORK

An ontology-based data integration system for geospatial data is proposed by Fonseca *et al.* [8, 9]. They introduce an Ontology-Driven Geographic Information System (ODGIS) that acts as a system integrator independent of the geographic data model. To build ODGIS, ontologies are specified using an ontology editor that stores formal representations of the ontologies. ODGIS provides a translation of the ontologies into an object-oriented framework, thus making ontologies available as classes, which are connected to geographic data sources through a mediator system. Ontologies in ODGIS are classified into levels, where the highest level describes the most basic concepts in the geospatial domain. Medium level ontologies contain more detailed concepts in the domain, which are derived from concepts in the higher level ontologies. To achieve data integration, Geospatial Information Communities (GIC) of common interest gather to define a higher level ontology from their own existing ontologies or if they do not have preexisting ontologies, they can define a high level ontology and use it to derive their own ontologies in the future.

Hovy proposes a domain independent semi-automatic ontology alignment approach for combining and standardizing large ontologies based on their lexical similarity [12]. Multiple languages are also supported by consulting dictionaries. Different kinds of matches, including text and hierarchy matches are used in the heuristics for the alignment process. Initially the unaligned ontologies are loaded and brought into a partially aligned state. For all unaligned concepts, the heuristics mentioned earlier are used to create a set of cross-ontology match scores. Then, a new set of alignment suggestions are created by combining the match scores using a function. The user can then check the suggestions and retain the best matches.

Chimaera is a software tool developed by the KSL group at Stanford, also used for merging and testing large ontologies [13]. It provides tools for merging different ontologies created by different authors and for diagnosing individual or multiple ontologies. Similarly to Hovy's approach, the tool allows users to check the automated merging procedure. Classes that need the user's attention are highlighted. PROMPT, which also uses a frame-based model for the ontology, supports merges at the slot level, in addition to merges at the class (frame) level [16].

In the ontologies that we consider for land use and wetland resource management, an ontology is a hierarchy where concepts are represented by the land use codes or the wetland components (the vertices in the hierarchy) and relationships are established between a parent concept and a child concept (the edges between the corresponding vertices). Such relationships represent generalization/specialization between the concepts. In our case, there are no explicitly represented properties or attributes associated with the codes. We have a simpler structure than that found in other systems [13, 16], and therefore the decision of whether two concepts match have to be solely based on the codes or component names.

Another system for the alignment of ontologies proposes a comprehensive approach encompassing a rich set of mapping types [10]. The authors stress the importance of declaratively specified mappings (our mappings are declarative), the necessity for automation, and of addressing the problem of defining a measure for the expressiveness of the mappings supported. However, several of these issues are left as future work.

The MOMIS approach has several similarities with our approach [1, 2]. This includes the semi-automatic processing of ontology alignment, but it has several different characteristics. For example, subsumption and lexical comparison are not possible in our case given that we have only strings to compare, not frames, and that lexical comparison, as that provided by WordNet [14], would only play an auxiliary role in the types of ontologies we currently consider.

The Clio project includes a research prototype of a tool that is used in schema mappings [11]. The user loads a source schema and a target schema and establishes connections between objects in both schemas graphically. Such connections are referred to as value correspondences, which express how an object or more in the source schema is transformed into a target value. Clio has a mapping engine that incrementally produces SQL queries that realize the mappings implied by the correspondences. Clio systematically ranks and generates mappings and helps the user in selecting between alternative mappings if needed. Our mapping tool also provides automatic help to the user by evaluating previous mappings to produce new ones (refer to the section on *Ontology Alignment*). In the *Conclusions* section, we discuss future plans to further automate our tool to enhance the experience of aligning ontologies.

In their survey paper, Rahm and Bernstein present a taxonomy that covers existing approaches for schema matching [17]. They give a detailed description of each approach distinguishing between schema-level and instance-level, element-level and structure-level, and language based and constraints based matching algorithms. The paper concentrates on the concept of a match operator that maps concepts and schemas and proposes the combination of several matching types. Although at present we only deal with one matching strategy, we are considering incorporating multiple matching algorithms in a layered approach in the future (refer to the *Conclusions* section).

GEOSPATIAL ONTOLOGIES

In this section, we present the classification schemes for the geospatial applications that we consider, namely the land use data for the state of Wisconsin and the wetland classification systems for the US and South Africa. To represent the corresponding ontologies we use our own XML DTD [6].

Land Use Data

For land use data, we consider the Wisconsin Land Information System (WLIS). A land use database system stores information about land parcels in XML format. Our land parcel data contains an identification number for the parcel (represented by the tag *lid*), the category of land usage under which it is classified *(lucode)*, the file containing the pertinent shape information (*shape_file*), and information about the owner of the parcel (*owner_id*).

Land use categories include *agriculture*, *commerce*, *industry*, *institutions*, and *residences*. Storing the land use codes of land parcels helps in better planning for township development, transportation, taxation, and so on. A typical query such as "*Where are all the crop and pasture lands in a region comprising Dane County (which contains the City of Madison), Racine County, and Eau Claire County?*" would be relatively straightforward when using one data set but more difficult when posed over a larger geographic area. Table 1 illustrates the heterogeneity of attribute names and values that would satisfy the criteria of the query over selected multiple data sets:

Planning Authority	Attribute	Code	Description
Dane County RPC	Lucode	91	Cropland Pasture
Racine County (SEWRPC)	Tag	811	Cropland
		815	Pasture and Other Agriculture
Eau Claire County	Lu1	AA	General Agriculture
City of Madison	Lu_4_4	8110	Farms

Table 1: Heterogeneity of attribute names and values.

1. *Lucode, Tag, Lu1* and *Lu_4_4* must be resolved as synonyms for the attribute that represents the land use code in the ontology.

2. Racine County data source uses two land use codes (811 and 815) for denoting croplands and pastures, while another jurisdiction uses one (Dane County RPC) and other jurisdictions are less specific (Eau Claire County and the city of Madison).

Wetland Classifications

Organizations monitoring the wetlands data inventory have an interest in sharing data. The lack of standard classification has long been identified as an obstacle to the development, implementation, and monitoring of wetland conservation strategies both at the national and regional levels [7]. In defining wetlands, the United States adopts the "Cowardin" Wetland Classification System [5] shown in Figure 3. In contrast, European nations use the International Ramsar Convention Definition (http://www.ramsar.org), and South Africa uses the National Wetland Classification Inventory [7] shown in Figures 4 and 5. Most classifications recognize the need for regionalization because of the variations in climate, geology, soils, and vegetation. Regionalization is designed to facilitate three activities: (i) planning, where it is necessary to study management problems and potential solutions on a regional basis; (ii) organization and retrieval of data gathered in a resource inventory; and (iii) interpretation of inventory data, including differences in indicator plants and animals among the regions. It can thus be concluded that it is extremely difficult to have a standardized classification system between nations and also between regions of a country with a large geographic area [5].

In this paper, we focus on the "Cowardin" Wetland Classification System and on the South African National Wetland Classification Inventory. Both classifications exhibit similar structure with some heterogeneities.
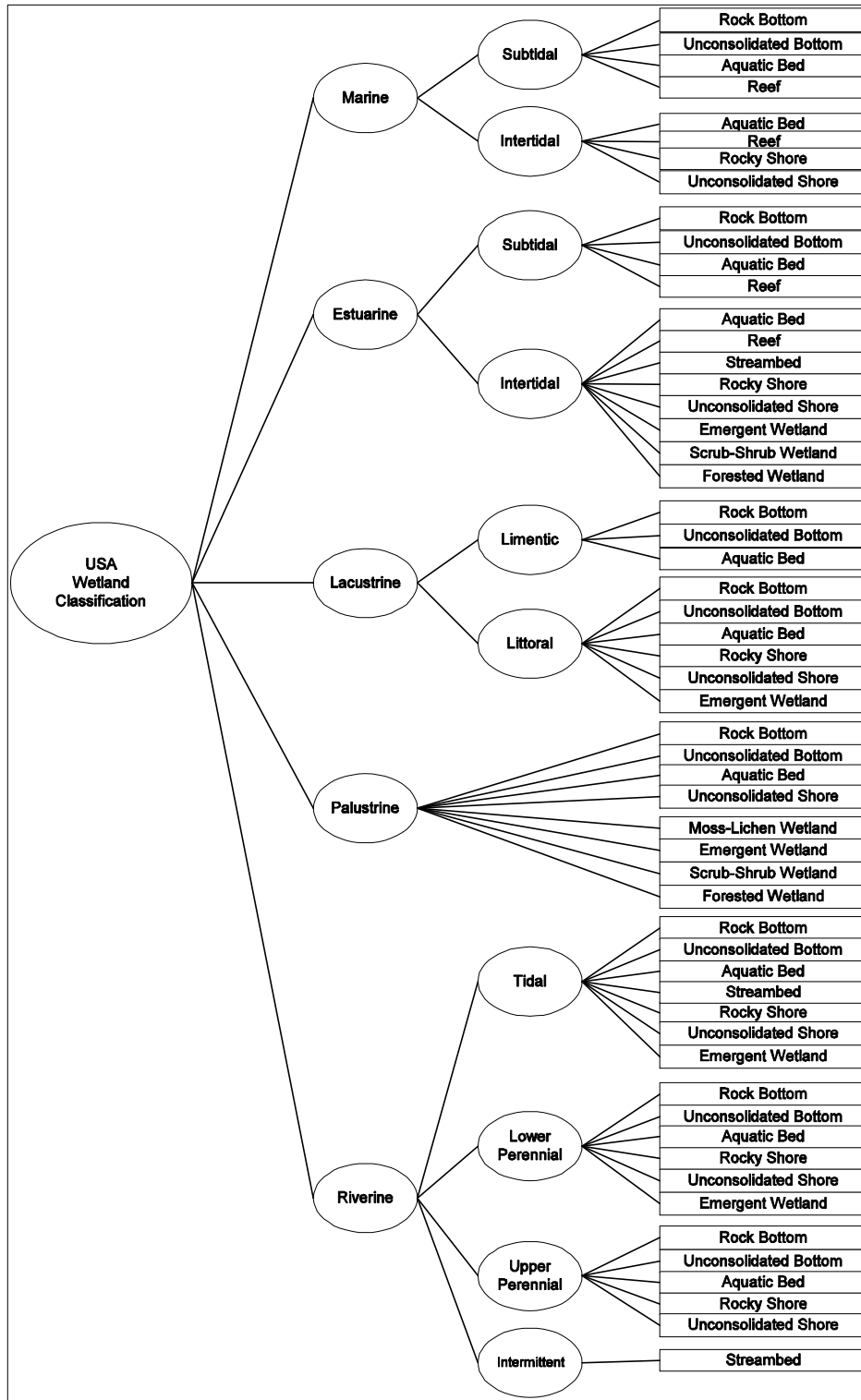
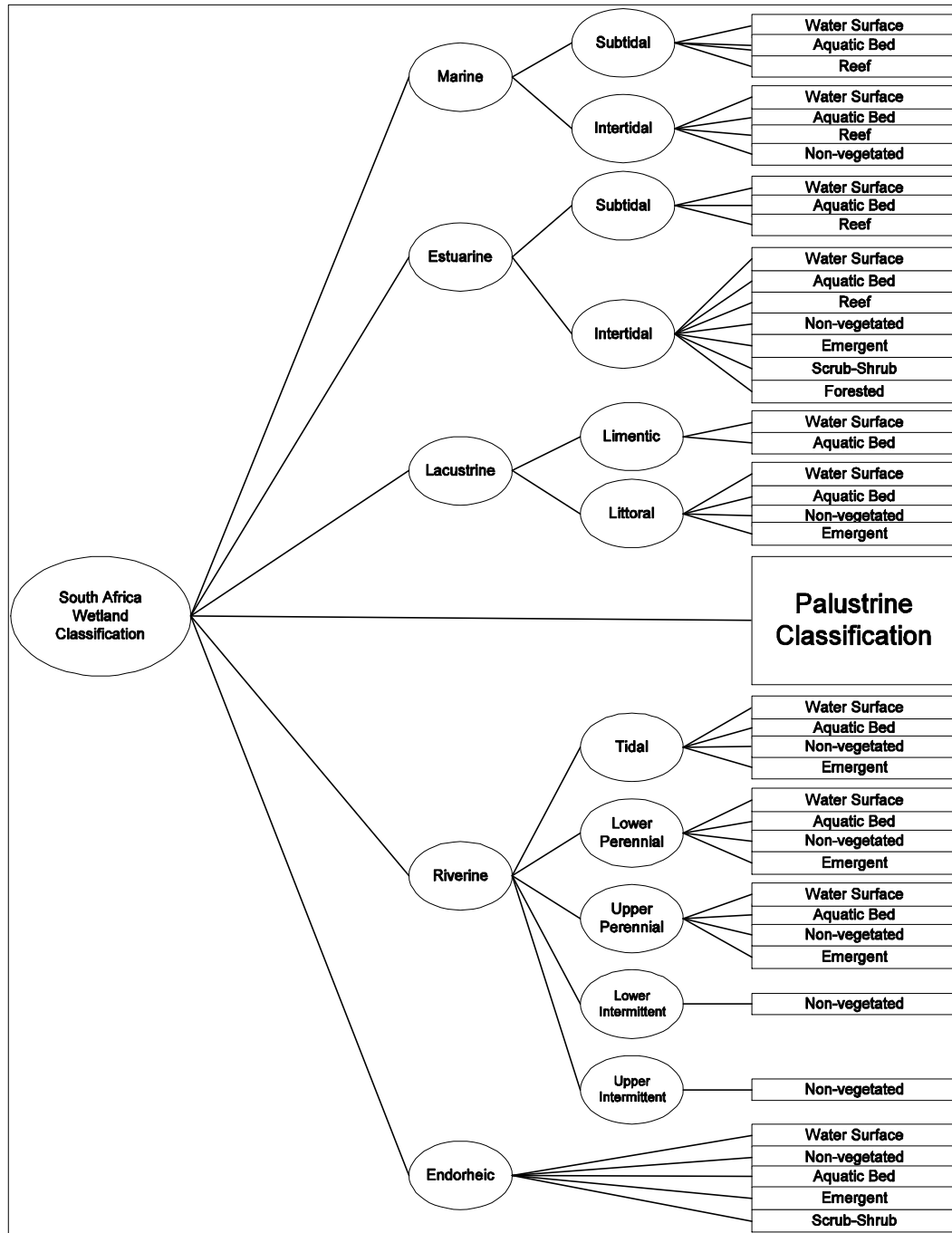Figure 3: Five "Cowardin" Wetland Classification System.

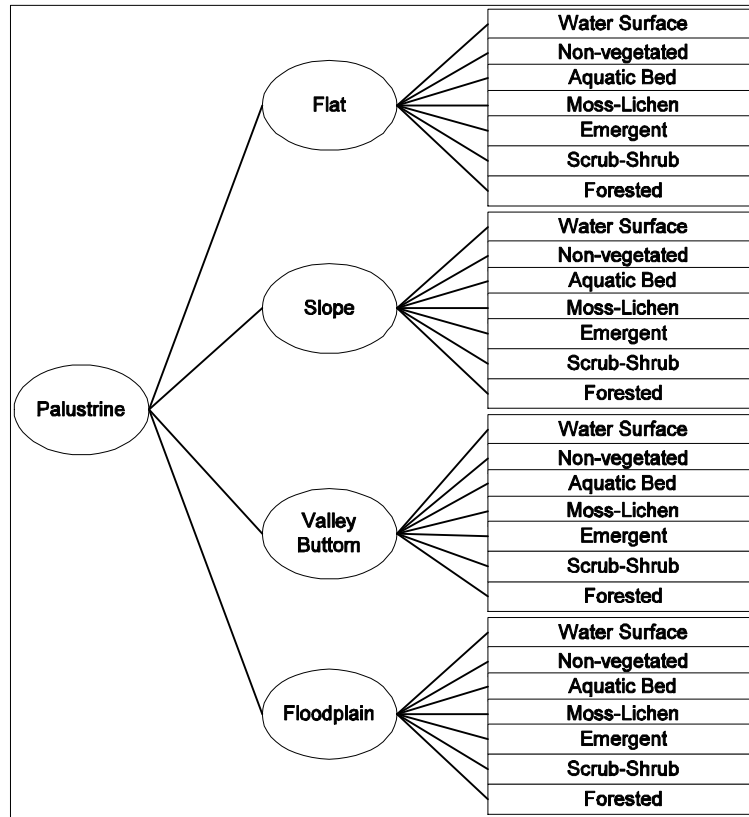Figure 4: South African National Wetland Inventory.

Figure 5: South African wetland palustrine classification.

"Cowardin" Wetland Classification System

Five systems form the highest level of the wetland classification hierarchy: *Marine*, *Estuarine*, *Riverine*, *Lacustrine*, and *Palustrine*. The *Marine* and *Estuarine* systems each have two subsystems called *Subtidal* and *Intertidal*. The *Riverine* system has four subsystems called *Tidal*, *Lower Perennial*, *Upper Perennial*, and *Intermittent*. The *Lacustrine* system has two subsystems, called *Littoral* and *Limnetic*. Finally, the Palustrine system has no subsystems. Within the subsystems, classes are based on the substrate material and flooding regime, or on the vegetative life form. The same classes may appear under one or more of the systems or subsystems. Six classes are based on the substrate and flooding regime: (i) *Rock Bottom*, (ii) *Unconsolidated Bottom*, (iii) *Rocky Shore*, (iv) *Unconsolidated Shore* with the same substrates as *Unconsolidated Bottom*, (v) *Streambed*, and (vi) *Reef*. The bottom classes, (i) and (ii) above, are flooded all or most of the time and the shore classes, (iii) and (iv), are exposed most of the time. The class *Streambed* is restricted to channels of intermittent streams and tidal channels that are dewatered at low tide. The life form of the dominant vegetation defines the five classes based on vegetative form: (i) *Aquatic Bed*, (ii) *Moss-Lichen Wetland*, (iii) *Emergent Wetland*, (iv) *Scrub-Shrub Wetland*, dominated by shrubs or small trees, and (v) *Forested Wetland*, dominated by large trees.

South African National Wetland Classification Inventory

The South African Wetland Classification Inventory is based on the system described above but has been regionalized to accommodate the full range of South African wetland diversity. For example, the riverine system has been divided into five subsystems, whereas the "Cowardin" system makes use of four. These subsystems will accommodate wetlands in channels containing flowing water for only part of the year. With much of South Africa experiencing a semi-arid climate, non-perennial riverine ecosystems are common. Two other modifications have been introduced in the South Africa classification [7]. First, a new system called Endorheic System has been introduced to identify small

closed basins referred to as "Pans" in South Africa which have a significant ecological role in that country. Secondly, the South African classification creates four subsystems under the Palustrine System based on the position of the wetland in the landscape.

ONTOLOGY ALIGNMENT

Given different ontologies, *ontology alignment* is the process of identifying related concepts that match in definition and purpose of use (within the context of the domain); the purpose is to make the ontologies coherent and consistent with each other. While establishing semantic relationship between concepts in different local ontologies can be challenging, the thorough identification of such relationships is essential for the development of accurate machine-based techniques to handle them. Past efforts in data integration used queries to map concepts and translate data instances between different schemas [15]. However, the specification of mappings using queries is an implicit rather than explicit method for representing the semantics between two ontologies. On the other hand, the explicit representation of the mappings can simplify the mapping process, the maintenance of the mappings, and enable users to better understand the different concepts and their correspondences.

In our examples, we represent the ontologies as trees. Each tree contains vertices that can be potentially mapped to vertices in another tree. For example, Figure 6 shows two subtrees, the left subtree is taken from the South African Classification (Figure 4) and the right subtree is taken from the "Cowardin" Classification System (Figure 3). The vertex *Aquatic Bed* in the left subtree is mapped to the vertex *Aquatic Bed* in the right subtree. The vertex *Water Surface* in the left subtree is mapped to both *Unconsolidated Bottom* and *Rock Bottom* vertices because it contains both of them. We also notice that *Limentic* vertex in the left subtree is mapped to *Limentic* in the right subtree.

Mapping Types

In Figure 6, *Aquatic Bed* in the South African classification and *Aquatic Bed* in the "Cowardin" classification are semantically at the same level of detail. Similarly, the vertices *Water Surface* and *Unconsolidated Bottom, Rock Bottom* are also at the same level. We say that such concepts are *aligned*. Initially, the information as to which concepts in the different ontologies are aligned must be provided by the user, that is, the ontology expert. Once two concepts are known to be aligned, the nature of the relation between them can be characterized using the following mapping types: *Exact*, the connected vertices are semantically equivalent in definition and purpose of use; *Approximate*, the connected vertices are semantically approximate in meaning and purpose of use; *Null*, the vertex in a particular ontology does not have a semantically related vertex in the other ontology with which it can be matched; *Superset*, A vertex in one ontology is more general in meaning than a vertex or a set of vertices together in the other ontology; and *Subset*, the vertex or a set of vertices together in one ontology are less general in meaning than the vertex in the other ontology. In Figure 6, *Aquatic Bed* in the left subtree should be mapped to *Aquatic Bed* in the right subtree using the mapping type *exact*. Similarly, *Water Surface* in the left subtree should be mapped to both *Unconsolidated Bottom* and *Rock Bottom* using the mapping type *Superset*. In this paper, we assume that the direction of the mappings in every figure is from the left ontological tree to the right ontological tree. Figure 7 illustrates several mappings types between vertices in two ontologies for land use patterns in a centralized integrated system. We show the global ontology subtree on the left-hand side of the figure and the local ontology subtree on the right-hand side. The vertices corresponding to *Industry*, *Mining* and *Manufacturing* in the global ontology can be mapped respectively to the vertices *Industrial Sector*, *Mining* and *Production* in the local ontology. In the global ontology, the vertex *Plastic wares* denotes concepts that are made of plastic or glass. However, in the local ontology, there is a vertex *Plastics* and another vertex *Rubber and Glass*, which denotes manufactured objects made of rubber or glass.
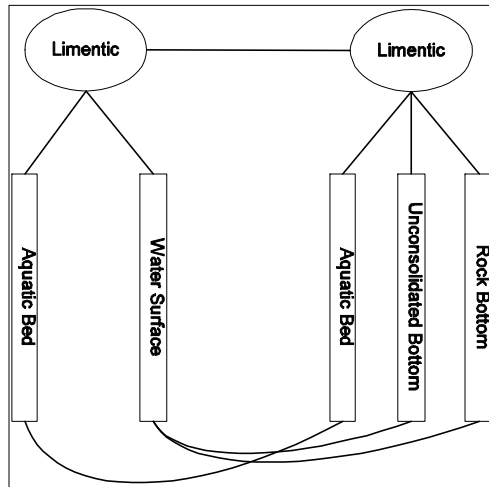
Figure 6: South African wetland palustrine classification.

The *Manufacturing* and the *Production* vertices are aligned. Similarly, the two *Mining* vertices are also aligned. *Manufacturing* is equivalent in meaning to *Production*, as both denote a collection of industries producing plastics, glass, and rubber products. Hence, this mapping is of type *exact* as denoted in the mapping from the *Manufacturing* vertex to the *Production* vertex. *Plastic wares* is a *Superset* of the *Plastics* vertex and *Rubber* is a *Subset* of the *Rubber and Glass* vertex.
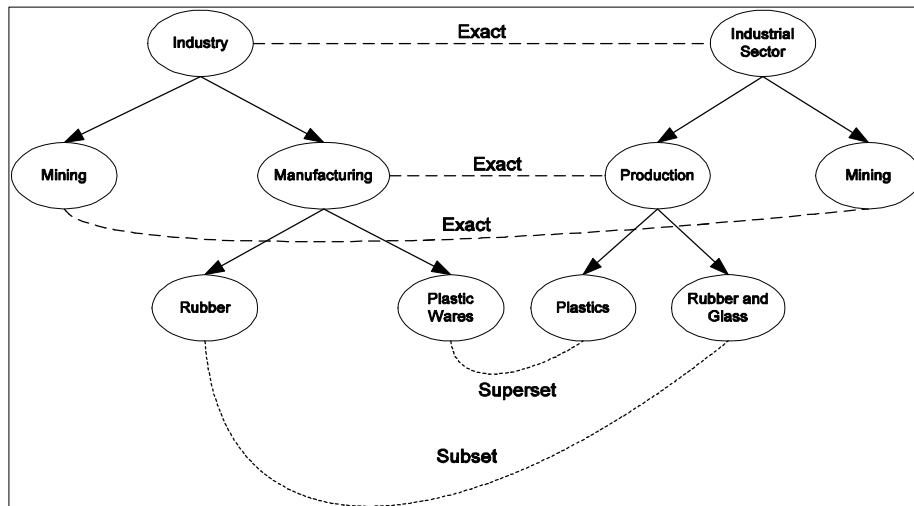

Figure 7: Mapping types.

Semi-automatic Alignment

To allow for the semi-automatic processing of the alignment of ontologies, we propose a framework that defines the values associated with the vertices of the ontology in two possible ways: as functions of the values of the children vertices or of the user input. We need to establish two assumptions to guarantee the correctness of the deduction process. The first one is that the specialization of a vertex in the ontology must be total, that is, each higher-level concept must belong to a lower-level concept in the hierarchy. The second one is that "bowties" [12], which are inversions in the order of the two ontologies that are being aligned, do not occur.

Our mapping techniques can be integrated in a semi-automatic alignment methodology to simplify the task of aligning large ontologies [12] and to aid the user in making correct decisions in the

mapping process. The user initially identifies the hierarchy levels in the two ontologies that are aligned. Then the alignment component propagates. When ambiguities or inconsistencies are encountered, or the algorithm cannot propagate values any further, those vertices are singled out. The user can then manually assist the algorithm by mapping concepts by hand similarly to other approaches [13, 16].

For example, in Figure 8, vertices *b* and *c* in the ontology on the left are mapped using mapping types *Exact* and *Superset* to vertices *e* and *f* in the ontology on the right. The mapping type between their parents *a* and *d* can be deduced to be *Superset* based on the mapping between the children, because we consider that the semantic content of the parent is the aggregation of the semantic contents of its children. All the children of *d* are mapped to children of *a*. This is the *Fully Mapped (FM)* case.
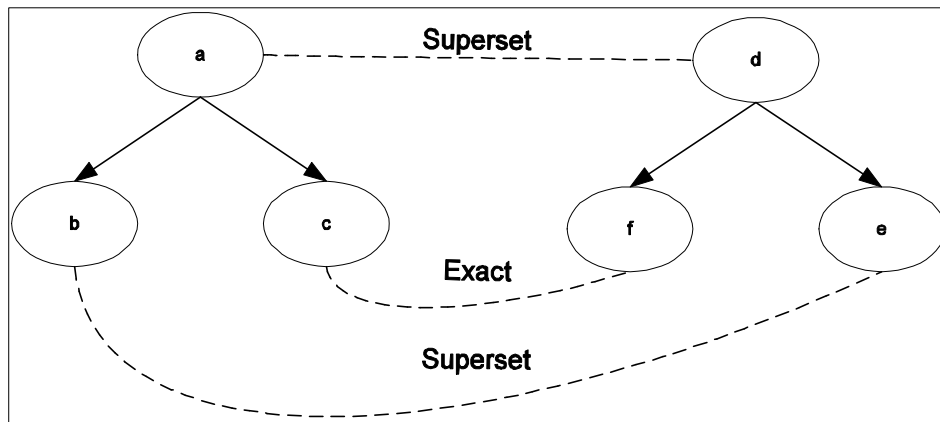

Figure 8: Fully mapped deduction operation.

The *Partially Mapped (PM)* case occurs if there are some children in one ontology that cannot be mapped to any of the children in the other ontology. For example, in Figure 9 vertices *b* and *c* in the ontology on the left are mapped to vertices *e* and *f* using mapping type *Exact*. But vertex *g* cannot be mapped to any of the vertices in the ontology on the left. As a result, vertex *a* is mapped to vertex *d* using mapping type *Subset*.
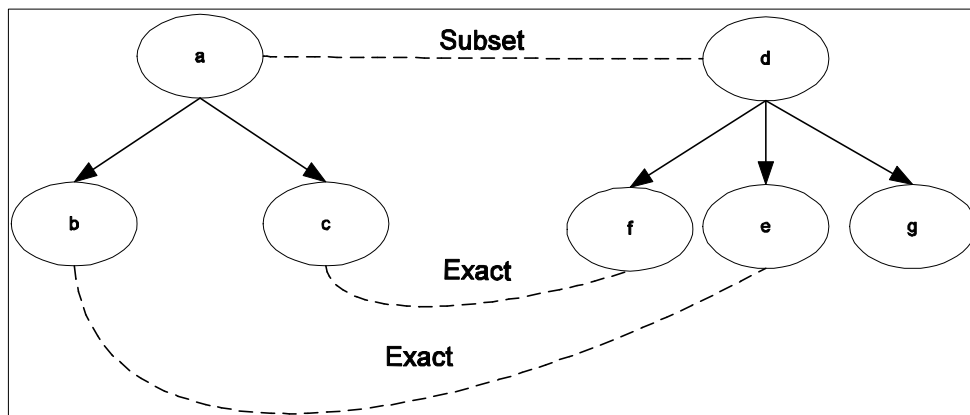

Figure 9: Partially mapped deduction operation.

Table 2 lists the different possible combinations of vertex mappings and the resulting mapping for

their parents. The *Fully Mapped (FM)* and *Partially Mapped (PM)* cases are shown in columns 3 and 4 of the table. A *User-defined* entry in the table indicates that the parent's mapping type cannot be automatically deduced and the user has to provide the appropriate mapping type by hand. These deduction operations can easily scale up to include the cases where a vertex has more than two children. They will be performed recursively, starting from the vertices that are aligned and traveling up the ontology tree, to deduce the mapping types between the ontologies being aligned. Any of the deduced mappings can be overridden by the user.

Child 1	Child 2	FM	PM
Exact	Exact	Exact	Subset
Exact	Approximate	Approximate	Subset
Exact	Superset	Superset	User-defined
Exact	Subset	Subset	Subset
Exact	Null	Superset	User-defined
Approximate	Approximate	Approximate	Subset
Approximate	Superset	Superset	User-defined
Approximate	Subset	Subset	Subset
Approximate	Null	Superset	User-defined
Superset	Superset	Superset	User-defined
Superset	Subset	User-defined	User-defined
Superset	Null	Superset	User-defined
Subset	Subset	Subset	User-defined
Subset	Null	User-defined	User-defined
Null	Null	User-defined	User-defined

Table 2: Automatic mapping deduction operations.

AGREEMENT MAKER

User Interaction

The *Agreement Maker* is a software tool that is used to create the mappings between two different ontologies and generate an agreement document, which is used by the query processor [6]. The tool implements our automatic deduction algorithm with which the user can invoke the deduction process on demand. Figure 10 shows the user interface.
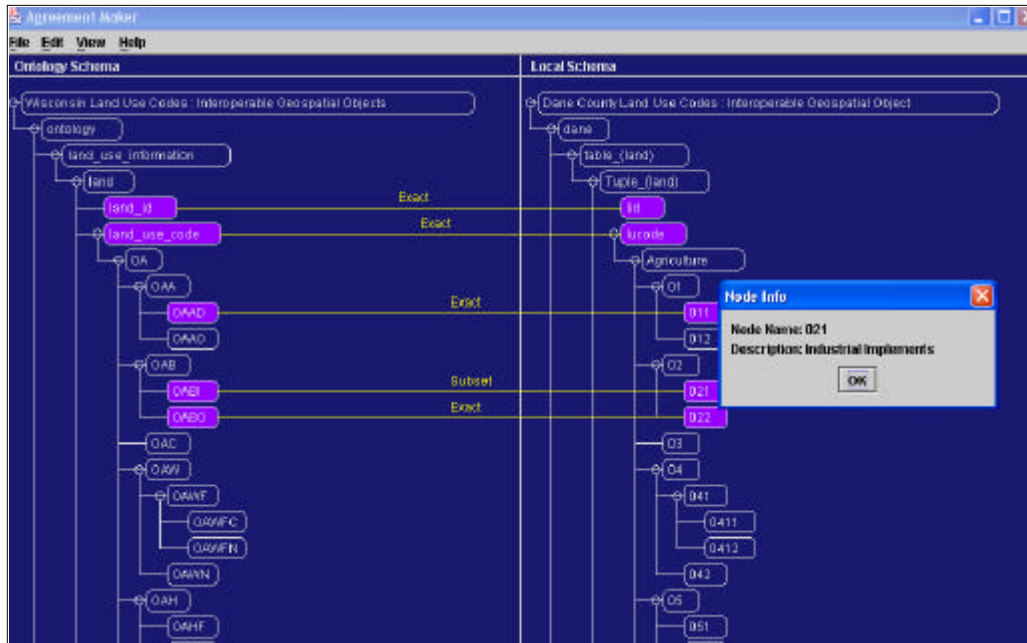
Figure 10: User interface showing established mappings.

In a centralized integrated system, an ontology expert establishes mappings between the global ontology and a local ontology. In a peer-to-peer integrated system, the expert maps the ontologies associated with two peers. The agreement maker shows the two hierarchies that represent the ontologies in two separate panes, enabling the expert to browse through the contents of each of the ontologies and establish the mappings between a concept (or concepts) in one of the ontologies and a concept (or concepts) in the other ontology.

The expert selects related concepts in both ontologies and chooses the mapping type that best explains the relation between the concepts, from the mapping types that were introduced earlier. Mapped concepts are displayed with an annotated line (showing the mapping type) connecting them to each other, as shown in Figure 10.

The agreement document is represented using XML, as part of our declarative approach to the integration of data sources [6]. In the agreement document, each concept in the first ontology that is being aligned is stored as the value of attribute *attrname id.* The corresponding concept (or concepts) in the other ontology is (are) stored as the value of attribute *equiv.* The mapping type is also stored as the value of attribute *mapping.* For example, in the alignment process shown in Figure 10 for the land use code integration scenario, the correspondences between the *land_use_code* and *lucode* and between *OAAD (Dairying)* and *011 (Dairying)* are specified in the agreement document as follows:

<attrname id= "land_use_code" mapping= "exact" equiv= "lucode">

<attename id= "OAAD" mapping= "exact" equiv= "011">

Deduction Module

The semi-automatic alignment methodology discussed earlier is integrated into the Agreement Maker system. Once the leaves of the hierarchy (or the vertices at a certain level of the two hierarchies) are mapped by the user, the mapping between the parents of the vertices on both ontologies can be inferred by the automatic mapping deduction operations of Table 2. In the user interface, the color of the lines that connect related concepts that were automatically deduced is made different from

that of the lines generated manually so as to be easily distinguishable.
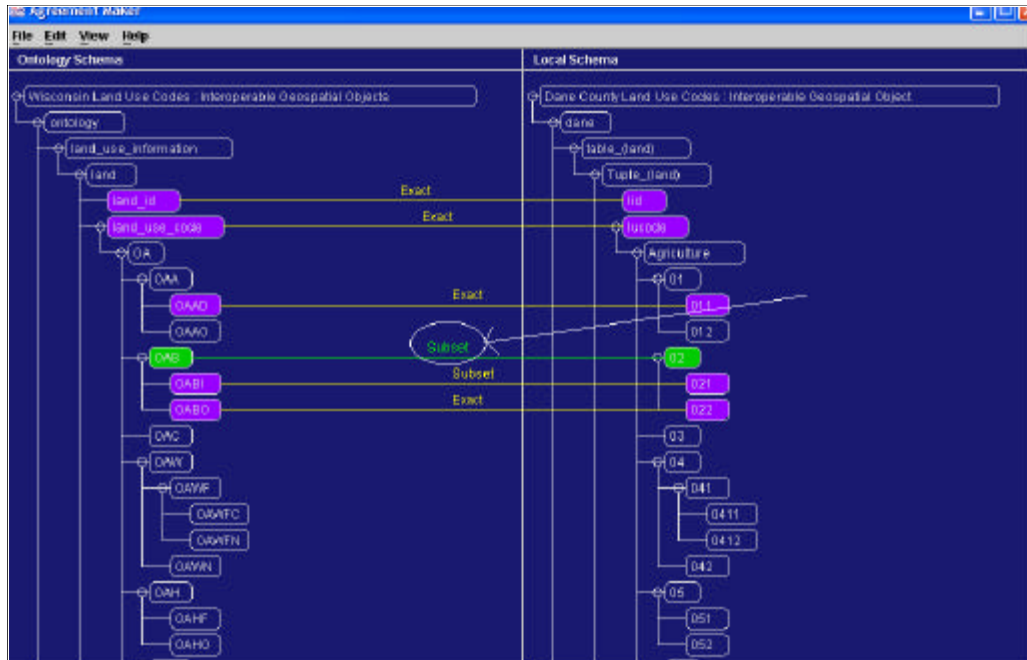

Figure 11: The mappings after running the deduction process.

The decision of the user overrides any mappings by the automatic deduction mechanism. Also, if the user has already mapped a vertex, no deductions will be considered for that vertex.

Figure 11 shows the vertices that were mapped and the mapping types. For example, the land use code *OABI (Industrial Implements)* in the global ontology is mapped by the user to the land use code *021 (Industrial Implements)* in the local ontology using the mapping type *Subset.* Also, the figure shows that the land use code *(Building)* in the global ontology is mapped to *02 (Building)* in the local ontology using the mapping type *Subset* by the deduction mechanism (pointed to by an arrow in the figure).

EXPERIMENTAL RESULTS

In this section, we perform an experiment with the Agreement Maker tool in the context of a peer-to-peer integrated system. Our objective is to illustrate the deduction process and to quantify the improvement brought by the use of the deduction process.

For our experiment we chose to align the "Cowardin" Wetland Classification System shown in Figure 3 with the South African National Wetland Classification Inventory shown in Figure 4. We performed the mappings in cycles, where each cycle consists of two phases: (i) manual mapping by the user and (ii) automatic mapping performed by the system when the deduction process is invoked. The next cycle starts after the second phase of the previous cycle ends.
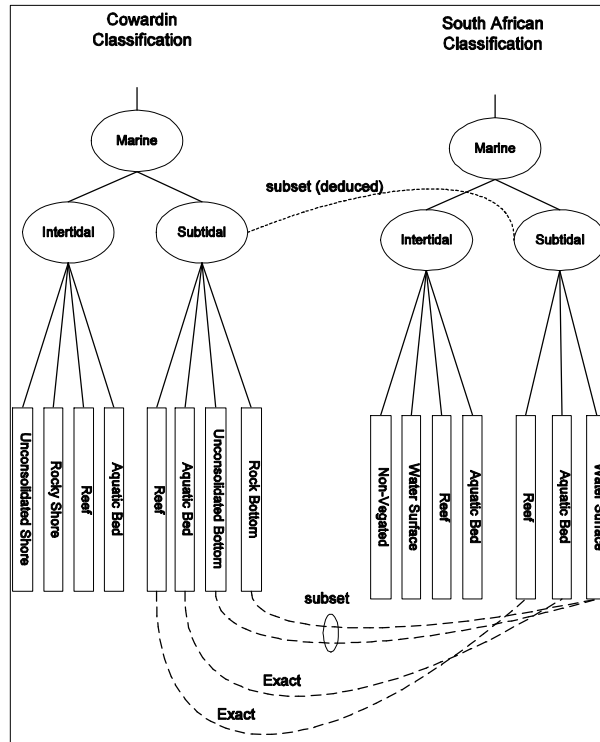
Figure 12: The first mapping cycle.

Figure 12 shows the detailed mappings of the first cycle. The user starts by mapping *Rock Bottom* and *Unconsolidated Bottom* in the "Cowardin" ontology to *Water Surface* in the South African ontology using *Subset* as the mapping type. Then the user maps *Aquatic Bed* in the "Cowardin" ontology to *Aquatic Bed* in the South African ontology using *Exact* as the mapping type. After this, the user maps *Reef* in the "Cowardin" ontology to *Reef* in the South African using *Exact* as the mapping type. This action ends the first phase of the cycle, and the user invokes the automatic deduction process, which will deduce that *Subtidal* in the "Cowardin" ontology is mapped to *Subtidal* in the South African ontology using the mapping type *Subset* based on the deduction rules in Table 2. Another cycle is then performed. We show the complete statistics of our experiment in Table 3 where we display the following values:

1. **MBU**: Number of mappings performed by the user in the first phase of a mapping cycle.

2. **MBS**: Number of automatic mappings performed by the system in the second phase of a mapping cycle.

3. **DR**: Deduction ratio, that is, the ratio of the number of mappings performed by the system to the number of mappings performed by the user in a mapping cycle.

4. **CDR**: Cumulative deduction ratio, that is, the ratio of the number of mappings performed by the system to the number of mappings performed by the user from the first to the current mapping cycle.

Mapping Cycle	MBU	MBS	DR	CDR
1	3	1	1/3	1/3
2	3	0	0/3	1/6
3	1	1	1/1	2/7
4	3	1	1/3	3/10
5	6	0	0/6	3/16
6	1	1	1/1	4/17
7	4	0	0/4	4/21
8	1	0	0/1	4/22
9	4	1	1/4	5/26
10	3	1	1/3	6/29
11	1	0	0/1	6/30
12	1	0	0/1	6/31
13	2	1	1/2	7/33
14	4	2	2/4	9/37
15	7	1	1/7	10/44
Total	44	10		

Table 3: Experimental results.

In this experiment, a total of 54 mappings were performed of which 44 were performed manually by the user and 10 were automatically deduced by the system. The percentage of mappings deduced by the system is influenced by the depth of the ontology trees and the number of the leaf nodes. In general, the leaf vertices are manually mapped by the user leaving the higher level vertices to be potentially mapped by the tool. In our experiment, both trees were shallow and wide and therefore only 18% of the mappings were deduced by the system. We anticipate that the deeper the ontology tree is and the fewer the number of the leaf nodes is, the higher the cumulative deduction ratio will be, as the number of possible deductions will increase.

CONCLUSIONS

In this paper, we describe an ontology alignment approach for the integration of semantically heterogeneous data with applications to two scenarios in the geospatial domain. In particular, we considered the alignment of land use code ontologies in a centralized integrated system and of wetland ontologies in a peer-to-peer integrated system. In the centralized integrated system, we align each of the local ontologies with the global ontology while in the peer-to-peer integrated system, we considered the alignment of the ontologies of two of the peers.

To facilitate the task of the user in the alignment process, we introduced semi-automatic deduction methods in a tool that supports the alignment of ontologies. We conducted an experiment in which we aligned two ontologies from the wetland classification domain. In this experiment 18% of the

mappings were deduced by the tool.

We are currently investigating a layered mapping approach where the mapping of concepts is done using multiple matching criteria. In our approach, the end user will select which criteria are relevant. Another issue is that we need to check the completeness of the mapping types we currently consider. We are currently introducing more mapping types to capture all possible relations in the different mapping layers.

From an application standpoint, we are also investigating ways to extend our alignment tool to encompass other applications such as in medical information systems.

REFERENCES

1. Benetti H., Beneventano D., Bergamaschi S., Guerra F., and Vincini M., An Information Integration Framework for E-Commerce. IEEE Intelligent Systems, 17:18-25, 2002.

2. Bergamaschi S., Guerra F., and Vincini M., A Data Integration Framework for E-Commerce Product Classification. In 1st International Semantic Web Conference (ISWC), pages 379-393, 2002.

3. Bishr Y., Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 12:299-314, 1998.

4. Calì A., Calvanese D., Giacomo G. D., and Lenzerini M., Accessing Data Integration Systems through Conceptual Schemas. In 20th International Conference on on Conceptual Modeling - ER, pages 270-284, 2001.

5. Cowardin L. M., Carter V., Golet F. C., and LaRoe E. T., Classification of Wetlands and Deepwater Habitats of the United States, 1979. http://www.npwrc.usgs.gov/resource/1998/classwet/classwet.htm.

6. Cruz I. F., Rajendran A., Sunna W., and Wiegand N., Handling Semantic Heterogeneities using Declarative Agreements. In International ACM GIS Symposium, pages 168-174, 2002.

7. Dini J., Cowan G., and Goodman P., South African National Wetland Inventory, 1998. http://www.ngo.grida.no.

8. Fonseca F., Egenhofer M., Agouris P., and Camara G., Using ontologies for integrated geographic information systems, GIS 6 (3): 231-257, 2002.

9. Fonseca F. T. and Egenhofer M. J., Ontology-driven geographic information systems. In ACM-GIS, pages 14-19, 1999.

10. Gennari J., Musen M., and Park J., Mappings for Reuse in Knowledge-based Systems. In 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW 98, 1998.

11. Hernandez M. A., Miller R. J., Haas L. M., Yan L., Ho C. T. H., and Tian X., Clio: A Semi-Automatic Tool for Schema Mapping. In SIGMOD Record, page 607, 2001.

12. Hovy E., Combining and Standardizing Large-Scale, Practical Ontologies for Machine Translation and Other Uses. In First International Conference on Languages Resources and Evaluation (LREC), 1998.

13. McGuinness D. L., Fikes R., Rice J., and Wilder S., An Environment for Merging and Testing Large Ontologies. In Seventeenth International Conference on Principles of Knowledge Representation and Reasoning (KR-2000), pages 483-493, 2000.

14. Miller G. A., WordNet: An Online Lexical Database. Technical report, Princeton University, 1990.

15. Miller R., Hernandez M. A., Haas L. M., Yan L., Ho C. H., Fagin R., and Popa L., The Clio Project: Managing Heterogeneity. ACM SIGMOD Record, pages 78-83, 2001.

16. Noy N. F. and Musen M. A., PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. In The Sixteenth National Conference on Artificial Intelligence (AAAI), pages 450-455, 2000.

17. Rahm E. and Bernstein P. A., A survey of approaches to automatic schema matching. VLDB Journal: Very Large Data Bases, 10(4):334-350, 2001.

AUTHORS INFORMATION

Isabel F. CRUZ
ifcs@cs.uic.edu
Department of Computer Science
University of Illinois at Chicago

William G. SUNNA
wsunna@cs.uic.edu
Department of Computer Science
University of Illinois at Chicago

Kalyan AYLOO
kayloo@cs.uic.edu
Department of Computer Science
University of Illinois at Chicago