Geographic Data Access in an Ontology-based Peer Data Management System

Rafael Figueiredo¹, Daniela Pitta¹, Ana Carolina Salgado², Damires Souza¹

¹ Federal Institute of Education, Science and Technology of Paraíba (IFPB), Brazil
{rafa.felype, daniela.pdj}@gmail.com, {damires}@ifpb.edu.br
² Center for Informatics (CIn), Federal University of Pernambuco (UFPE), Brazil
{acs}@cin.ufpe.br

Abstract. In distributed data environments, ontologies have been used as a support for data management. For instance, ontologies may be used to describe the semantics of data at different sources, helping to overcome problems of data heterogeneity and semantic interoperability. Generally, the task of accessing data by means of conceptual ontologies has been called Ontology-based Data Access (OBDA). A typical scenario for OBDA instantiation is a Peer Data Management System (PDMS) where queries submitted at a peer are answered with data residing at that peer and with data acquired from neighbor peers through the use of mappings. In this work, we apply the principles underlying an OBDA in the light of a PDMS, using geographic databases as data sources. When dealing with geospatial data, specific problems regarding query answering and data visualization may occur. To help matters, we propose an approach named easeGO, which provides access to a geographic database using an ontology. We also present a tool, which allows users to formulate queries using the peer ontology as well as visual elements and spatial operators. We present the principles underlying our approach, examples illustrating how it works and some obtained experimental results.

Categories and Subject Descriptors: H.2 [Database Management]: Miscellaneous; H.3 [Information Storage and Retrieval]: Miscellaneous; I.7 [Document and Text Processing]: Miscellaneous

Keywords: OBDA, PDMS, Query Answering, Query Reformulation

1. INTRODUCTION

In distributed data environments, particularly those involving data integration, ontologies have been formally used to describe the semantics of the data sources. The goal is both to facilitate the standardization using a common representation model, and the discovery of the sources that provide the desired information [Lopes et al. 2012]. The use of ontologies as a layer between the user and the data source (in this work, a geographic database) adds a conceptual level over the data. It allows the user to query the system using semantic concepts without taking care about specific information from the database. Generally, this type of access has been called Ontology-based Data Access (OBDA) [Calvanese et al. 2009] and its principles can be applied to any setting where query answering is accomplished using the ontologies that describe the sources. Typical scenarios for OBDA instantiation are Peer Data Management Systems (PDMS) [Souza et al. 2011], Data Spaces [Hedeler et al. 2009] and the Semantic Web [Calvanese et al. 2009].

We apply the OBDA principles in a PDMS named SPEED - Semantic PEEr Data Management System [Pires et al. 2012]. The SPEED system is composed by data sources (called peers) and adopts an ontology-based approach to assist relevant issues in data management such as query answering. Query answering in SPEED means to provide capabilities of answering a query considering that such query is submitted over one of the peers and there is a set of mappings between the peer and their
neighbors. Particularly, in this work, we are using geographic databases as data sources. In order to uniformly deal with geospatial data without worrying about their specific heterogeneity restrictions (syntactic or semantic), we use ontologies as uniform conceptual representation of peer schemas. When a peer asks to enter the system, its schema is exported to a peer ontology. During the ontology building process, a set of correspondences (mappings) between the generated peer ontology components and the original database schema is also generated. We use the produced peer ontology and the set of correspondences to reformulate ontological queries into the database query language and retrieve corresponding instances from the geographic database.

One important issue in our work regards the use of geospatial data. A higher level of complexity is observed in geospatial data manipulation because of their special characteristics (e.g., spatial location, representation). Thus, there is also a need for special visualization tools and exploration mechanisms to make provision for the spatial presentation and querying of these data. Considering these presented aspects, our approach has been specified and developed. Named as Easy Geographical Ontological access (easeGO), it is concerned with two main issues: (i) an interface which allows working both with the peer ontology and a cartographic representation of the data (e.g., a map) to visualize the metadata and formulate queries and (ii) a query manager, which reformulates the query formulated in the interface (using the ontology or the map) into queries, which may be executed by the DBMS (e.g., in SQL). After executing the query, the query manager receives the results and represents their output according to the user preferences on data visualization. The easeGO interface has been designed following the principles of visual query systems (VQS) [Catarci et al. 1997]. In this light, it is based on using the peer ontology and on direct manipulation interaction mechanisms. It may be used by any user, including the ones who are not familiar with the syntax of query languages such as SQL or are not interested in learning a specific query language.

This article is organized as follows. Section 2 introduces the SPEED system. Section 3 presents the easeGO approach. Section 4 describes the developed easeGO tool with some accomplished experiments. Related work is discussed in Section 5. Finally, Section 6 draws our conclusions and points out some future work.

2. THE SPEED SYSTEM AS AN OBDA

Peer Data Management Systems (PDMS) are characterized by an architecture constituted by various autonomous and heterogeneous data sources (e.g., files, databases), here referred as peers. The SPEED (Semantic PEer Data Management System) system [Pires et al. 2012] is a PDMS that adopts an ontology-based approach to assist relevant issues in peer data management. Its architecture is based on clustering semantically similar peers in order to facilitate the establishment of semantic correspondences (mappings) between neighbor peers and, consequently, improve query answering. Peers are grouped according to their knowledge domain (e.g., Education, Tourism), forming semantic communities. Inside a community, peers are organized in a finer grouping level, named semantic clusters, where peers share similar ontologies (schemas). Particularly, in SPEED, peer ontologies are employed to represent the schema of the sources stored in peers. A peer has a module to translate an exported schema described in its original data model to an ontology representation.

In what follows, we discuss the ODBA principles and why we can consider SPEED as an OBDA. Also, we describe how the query reformulation process is accomplished in SPEED.

2.1 OBDA Principles

The paradigm of ontology-based data access (OBDA) has emerged as an alternative for assisting issues in data management (e.g., data sources heterogeneity), usually in distributed environments. The underlying idea is to facilitate access to data by separating the user from the data sources using an ontology [Kontchakov et al. 2011]. This ontology provides a user-oriented view of the data and
makes it accessible via queries formulated only in the ontology language without any knowledge of the data source schema [Calvanese et al. 2009]. OBDA settings have some common characteristics, such as [Lopes et al. 2012]: (i) the data sources usually exist independently of the ontologies which describe them, (ii) ontologies and data sources show diverse levels of abstraction and may be represented using different models; (iii) the ontology is the unique access point for the interaction between the users and the system; and (iv) queries submitted on the ontology must be answered using a set of existing mappings between the ontology elements and the data source schema.

Comparing PDMS and OBDA features, we can verify some common characteristics. A PDMS is a P2P system that provides users with an interface where queries are formulated transparently on heterogeneous and autonomous data sources [King et al. 2010]. The main service provided by a PDMS thus concerns query answering. Meanwhile, the main reason to build an OBDA system is to provide high-level interfaces (through ontologies) to the users of the system. In both settings, users should express their queries in terms of a data source view (i.e., an ontology), and the system should reformulate these submitted queries using existing mappings that help to translate them into suitable ones to be posed to the data sources.

Regarding these characteristics, and, since data sources schemas in SPEED are described using ontologies (named hereafter peer ontologies), we may consider the SPEED system as an OBDA setting.

2.2 Query Reformulation in SPEED

In SPEED, a query posed at a peer is routed to other peers to find answers to the query. An important step of this task is reformulating a query issued at a peer into a new query expressed in terms of a target peer, as depicted in Figure 1 (highlighted in a filled rectangle). To accomplish this task, a query reformulation module, named SemRef, has been developed [Souza et al. 2011]. The SemRef approach brings together both query enrichment and query reformulation techniques in order to provide users with a set of expanded answers. In this sense, it analyzes an initial query expression (posed in a source peer) in order to find out some extra semantic knowledge that can be added (in a target peer) at query reformulation time, so its resolution may provide expanded (additional related) answers. To this end, besides equivalence, it uses other correspondences which go beyond the ones commonly found (specialization, generalization), proposing disjointness and closeness, identified by a specific semantic matcher (called SemMatcher) [Souza et al. 2011].

Through the set of semantic correspondences, the SemRef approach produces two kinds of query reformulations: (i) an exact one, considering equivalence correspondences; and (ii) an enriched one, resulting from the set of the other correspondences. The priority is to produce the best query reformulation through equivalence correspondences, but if that is not possible, or if users define that it is relevant for them to receive semantically related answers, an enriched reformulation is also generated. Additionally, it takes into account the context of the user, of the query and of the environment as a way to enhance the process and to deal with information that can only be acquired on the fly. As a result, users are provided with a set of expanded answers, according to their preferences. In our previous works, however, the SemRef approach has taken into account only conventional data (i.e., data are not georeferenced).

In this work, the SPEED system has been instantiated with geographic databases. A tool named GeoMap was developed for automatically building a geospatial peer ontology [Lima et al. 2011]. This peer ontology represents a semantic view of metadata stored in a geographic database. During the mapping of these metadata, it is possible to identify equivalence correspondences between the generated ontology components and the existing database schema entities and properties. In order to define this set of equivalence correspondences, we build an OWL document composed by a specific construct named IsEquivalentTo. Such construct has been defined and used to indicate which ontology concept is equivalent to the database schema element. Also, it indicates which ontology properties are equivalent to the database schema attributes and relationships.
Thereby, query reformulation in SPEED can now be accomplished in two ways, as depicted in Figure 1: (i) vertically (highlighted in a filled rectangle), between a query submitted in a peer using its local ontology and the data source schema, and (ii) horizontally (highlighted in an unfilled rectangle), between a source and a target peer ontology (i.e., between two neighbor peers). The former is the focus of this work. Particularly, we are interested in the way we can use peer ontologies to formulate queries and execute them, retrieving real data from geographic databases.

3. THE EASEGO APPROACH

One of the most representative realms of diversity of data representation is the geospatial domain. Geospatial data, besides hierarchical and descriptive components (relationships and attributes), are featured by other ones such as geometry, geospatial location and capability of holding spatial relationships (e.g., topological) [Hess et al. 2007]. Furthermore, geospatial data are often described according to multiple perceptions, different terms and with different levels of detail. In our work, geospatial data are represented by means of the vector model. As a result, they are expressed as objects and are stored as points, lines or polygons, depending on the scale of their capture. In this sense, the syntactic, semantic and spatial data heterogeneity should be considered when dealing with geospatial data in a PDMS and in query answering processes.

On the other hand, an usual ontology is composed by concepts, properties, axioms and, optionally, instances. In order to deal with query reformulation, considering the vertical access shown in Figure 1, we have to deal with the correspondences between the peer ontology elements and their corresponding elements in the geographic database schema. The easeGO approach has been specified and developed to take into account the set of correspondences between the peer ontology and the geographic database schema elements, thus enabling query reformulation. Besides, the easeGO approach has put together two issues to facilitate query formulation by users who are unfamiliar with geospatial query languages: (i) visual query language concepts and (ii) OBDA principles. The former provides the user with visual elements that abstract the underlying query language syntax, helping to guide editing querying actions so as to minimize the risk of errors [Catarci et al. 1997]. As already mentioned, the latter provides a unique data access by means of an ontology (i.e., a peer ontology).

Considering these issues, the purpose of the presented easeGO approach is to support query formulation in the context of the SPEED system through a peer ontology and using geospatial visual elements. An overview of the easeGO architecture is depicted in Figure 2. In the following, we present its components which are divided into two main modules: (i) the interface, composed by data view and querying options and (ii) the query manager, responsible for reformulating the submitted queries and executing them, as well as formatting their results output.
3.1 The easeGO Interface: User Perspective

The first impression causes a very strong feeling, not just from person to person, but also between people and objects. This is also the case for computational system interfaces, especially those regarding the use of geospatial data. A geospatial data query interface design should deal with the characteristics and difficulties faced in the elaboration of a DBMS interface and provide the specific geographic application requirements, such as multiple representations for objects and spatial query formulation.

In this work, the interface design has the following goals: (i) users can be novices or experts, but our main purpose is to design an easy-to-use interface for the less experienced users, (ii) the interface should be capable of providing geospatial data exploration as well as making use of the peer ontology concepts to facilitate query formulation. Since we aim to provide geospatial query formulation, we have also to accommodate in the interface a way of manipulating spatial relationships (e.g., adjacency, cross) between entities that are geometrically defined and located in the geographic space. This process is accomplished by using visual elements to compose the query expression. Indeed, we try to apply the principles underlying the so-called Visual Query Systems (VQS) [Catarci et al. 1997]. VQS are characterized by features such as the use of icons and visual metaphors, instead of text, and the availability of interactive mechanisms to support query formulation.

The scenario in which we consider the deployment of our approach consists of a geographic database which provides its own query language (i.e., object-relational geographic databases). As shown in Figure 2, the easeGO interface adopts a hybrid strategy for formulating queries and is composed by the following options:

— View Ontology: the peer ontology, which describes a given geographic database, defines a vocabulary which is meant to be closer to the user’s vocabulary. The user can exploit the ontology concepts to formulate a query using search and navigational options. The ontology is depicted using tree or graph views.

— View Map: the geospatial data may be presented in a cartographic view using, for example, a map. This option gives the user a closer view of spatial reality where s/he is able to work with.

— Formulate Query: users may formulate queries using nodes and edges (which represent classes and properties) from the peer ontology. Each node/edge of the tree/graph corresponds to elements from the database schema. Once selected, a node becomes the focus for querying. Users may also formulate queries using visual elements provided by the map option. This option supports a predefined set of spatial operators that improves the easeGO query capability.

— View Results: users may define their preferences regarding the way they will see query results. The results may be shown using a table option (text data) or using the map, where resulting objects are highlighted in a different color.
When using the peer ontology to formulate a query, the user can select a node and request instances of this node. S/he may also, from this node, set the query in a visual way by using a form which is dynamically built. This form provides the existing properties of the chosen node. By using this form, the user chooses the properties s/he wants to view (as a project operation from the relational algebra) and determines the conditions (as a select operation from the relational algebra) that the query should verify. When formulating a query by using the map option, users may choose a geographic object to be a query operand and drag it to a query area. Once the user has selected the first query operand and it has been dragged to the query area, s/he selects the spatial operator to be applied. If it is a unary operation, the query may be validated. However, if it is a binary operation, another geographic object will be selected. From both query formulation options, a query Q (Figure 2) is generated. This query will be sent to the query manager, as explained in the following.

3.2 The easeGO Query Manager: Reformulating Queries

We define the query manager approach as follows: given a user query Q expressed in terms of the concepts of the peer ontology, a target geographic database schema GeoDB schema, and a set of correspondences between the peer ontology elements and the database schema ones, our goal is to find a reformulated query of Q expressed in terms of the concepts of the GeoDB schema in such a way that it may be executed by the DBMS. The reformulated query is named Q1 which is executed in the DBMS and the query results R1 are returned to the query manager. The query manager considers the user preferences regarding the data visualization and sets the resulting data R which is sent to the interface. R may be depicted using a table or highlighted on the map.

4. RESULTS AND EXPERIMENTS

We have implemented the easeGO approach by means of a querying tool in Java, using the OWL-Prefuse\(^1\) and GeoTools\(^2\) APIs. The tool provides access to geographic databases coded in Oracle Spatial\(^3\) and PostGIS\(^4\). When the user starts working, a peer ontology is depicted through a graph or tree representation. The peer ontology is related to a particular geographic database, which refers to a single geographic region. The user can navigate at the peer ontology level, browse geospatial data using layers over a map, or formulate queries. From the functional point of view, the easeGO tool current release provides the following aspects:

—Peer Ontology Navigation: the user is able to navigate over the ontology concepts and choose one for querying. This querying process may be accomplished in a two-fold way: (i) by retrieving all the instances of a given concept or, (ii) starting from a general concept, the user can choose the properties s/he wants to see and define constraints to be applied over the data.

—Form-based query formulation: after choosing a concept using the peer ontology, the tool provides the user with a form, which presents the concept properties and enables query constraints definition. Thus, s/he is able to fill in the form, by choosing the desired properties and establishing constraints, to create a query expression in a high-level way.

—Exploration of Geospatial Objects: it means that objects are shown in the visualization area and can be selected for getting information about their descriptive attributes, for visualization operations (zoom, pan) or for spatial queries. It is also possible to enable or disable object layers.

—Spatial Query Formulation: using the cartographic view, the process of building a query involves the following steps: (i) the geographic objects of interest are selected and dragged to a query building area, (ii) spatial operators are selected and (iii) the query is validated and then executed.

[^1]: http://owl2prefuse.sourceforge.net/
[^2]: http://www.geotools.org/
[^3]: http://www.oracle.com/
[^4]: http://postgis.refractions.net/
Fig. 3. Peer Ontology represented as a Graph and a Query Example

—Query Results Presentation: the tool may depict the query results by using a table with the answers or by highlighting the resulting geospatial objects on the cartographic view.

—Hints and help messages during the execution of each user task.

We provide some examples of these functionalities in the following.

4.1 easeGO in Practice

In the following examples, we use a geographic database with data about inhabitance control in Paraiba state (stored in PostGIS). Its scheme was previously mapped to a peer ontology representation. Initially, Figure 3 (option I) depicts the peer ontology by means of a graph structure. In this example, the user has selected the concept Usuario (which is highlighted) and a form-based query formulation option is presented to him/her (option II). This form is dynamically generated according to the underlying properties of the chosen ontology concept. The form shows the existing properties of the node and enables their setting for query answers presentation. Besides, the form lets the user to define constraints using the existing properties and commonly used operators (e.g., equality and logical operators). The user, then, fills in the form with his/her preferences and definitions. The tool generates a query which will be reformulated and executed. In this example, the user has chosen the concept Usuario, together with the following properties: usuariologin, usuarionome and usuarioemail. In addition, s/he has defined a condition over the user name (usuarionome = Gustavo Brasileiro). A fragment of the query results is also shown in Figure 3 (option III).

To allow geospatial objects exploration, the easeGO tool provides a cartographic view of the data, as shown in Figure 4. This view is composed by three main areas, as follows: (i) geospatial objects area, (ii) spatial operators area, which depicts the set of available spatial operators using icons (based on the standard operators provided by PostGIS), and (iii) a query formulation area, where a visual query may be built. In this case, when a geographic object of an active layer is selected, it is represented as an icon and may be dragged to the query area as a query operand. In Figure 4, objects belonging to the Area-cadastro layer are shown in the geospatial objects area. In this example, we show a visual query formulation where the user has selected a geographic object from the geospatial objects area (option I), together with the disjoint spatial operator (option II). The visual query is built in the
Fig. 4. Cartographic View of Data and a Visual Spatial Query Example

query area (option III) and its results are highlighted on the map (option IV).

4.2 Experiments

We have conducted some experiments to verify the effectiveness of our approach. The goal of our experiments is two-fold: (i) to check whether the user is able to formulate queries easily using the peer ontology and the geospatial visual elements and (ii) to verify how the query manager accomplishes the query reformulation process. To this end, we have invited some users to evaluate our tool. The user group was composed at most by undergraduate students in Computer Science, Engineering and professionals used with GIS functionalities. After explaining the goal of the evaluation, we let them interact with the tool for a few moments until they got used with it. Then they used the tool and received a questionnaire to be filled out.

Users filled out a questionnaire stating their opinions on the interface design, the use of peer ontologies and the map view, and the way query results were presented. Five measures were required: learning facility (in which degree the tool is easy to learn to use), query formulation facility (in which degree the user considers as an easy process to formulate a query), design issues (in which degree the interface layout contributes to query formulation and data visualization), results clarity (in which degree the answers were free of ambiguity), and results satisfaction (in which degree the answers fulfilled the required query).

Figure 5 presents a summary of the evaluation regarding the peer ontology access option. In terms of learning facility, query formulation, results clarity and satisfaction, the users provided a good or even great impression. Only some of them considered the interface layout hard to understand and suggested some improvements such as: a better way of presenting the query results, functions provided on the map option should be also available in the peer ontology view and the interface design could be better. Figure 5 also presents the users perceptions on the map access option. Since most of users were not used to deal with geospatial data and queries (i.e., only a few of them are GIS users), they had more difficulty to learn about how to use the map and to formulate queries. The main problem regarding query formulation was indeed the fact that they did not know the semantics underlying the spatial operators. Nevertheless, after learning the overall principles, they could then accomplish the task properly. Thus, after this initial impression, they were comfortable to formulate queries and clearly visualize the produced results. In this sense, the outcome of the experiments indicated that the tool can be used also by less-experienced users to query a domain (in this case, a geographic one).
5. RELATED WORK

Currently there are many approaches that make use of query interfaces using ontologies. As an example, the Quelo system [Franconi et al. 2010] allows access to heterogeneous data sources through a visual interface and its reasoning processes are driven by an ontology. Particularly in the geospatial realm, Grandchamp [2011] presents the specifications for a conceptual layer in GIS as an intermediate step to solve the matching problem of two corresponding objects or boundaries during the combination of different information layers. The work of Zhao and his group [Zhao et al. 2008] provides integrated access to distributed geospatial data using RDF ontology and query rewriting. Zhai et al. [2010] use SPARQL to perform semantic query and retrieval of geospatial information which has been converted (metadata and instances) to a geospatial ontology described by OWL. Baglioni et al. [2008] create an intermediate semantic layer between the user and the geodatabase in order to facilitate the users queries. They also enrich the generated ontology with semantics from a domain ontology by finding correspondences between the classes and properties of the two ontologies. This work is the most similar to ours.

Recently, works aiming to make geospatial data accessible on the web have also been accomplished. The OnGIS system introduces an abstraction layer based on OWL ontologies [Smid and Kouba 2013]. It provides an intuitive searching GUI for exploring geospatial data semantically integrated from different sources, aiming to provide non-expert users with a simple querying interface. Another example regarding the use of web services to enable geographic data access is the work of He et al. [He et al. 2013], which presents a heterogeneous web service integration method based on dynamic semantic schema matching and automatic information retrieval and fusion. In the setting of linked data, the GeoKnow project makes OpenStreetMap data available as an RDF knowledge base [Garcia-Rojas et al. 2013]. As a result, OSM data were introduced in the LOD (Linked Open Data) cloud and interlinked with some other datasets.

Comparing these works with ours, we go one step further as we put together both OBDA principles and VQS ones using a peer ontology and visual elements to allow access over the geospatial data in the light of a PDMS setting. Another difference is the use of the correspondences set for allowing query reformulation.

6. CONSIDERATIONS AND FUTURE WORK

This work is an attempt to combine OBDA and PDMS principles as a way to enable the access to geospatial data. To achieve this, aspects related to geographic databases, query interface design, ontology navigation and query reformulation have been considered. The easeGO approach provides an intuitive and transparent setting where the user is able to work with a peer ontology or with a cartographic view of the geospatial data. A query formulated in the tool interface is reformulated by
the query manager using a set of existing correspondences between the peer ontology and the database schema. Query results may be visualized both in a table form or using the map.

Experiments accomplished with real users showed that the easeGO approach has some advantages: (i) it does not require users to have previous knowledge about the underlying database schema or query language; (ii) it gives the user a closer view of spatial reality where s/he is able to work with; (iii) it supports a predefined set of spatial operators that improves query capability and (iv) it allows users to pose queries by a visual, form or ontological paradigm, helped by tips that cover all tasks.

The easeGO approach has been implemented in the light of the SPEED PDMS system, although it can be applied to any OBDA environment which deals with geographic databases. As future work, it will be extended to provide query reformulation between two neighbor peers, taking into account the semantic correspondences between them.

REFERENCES


