

GeoNote: A Web Service for Geographic Data Annotation in Biodiversity Information Systems

Fabiana B. Gil^{1,2}, Nádia P. Kozievitch², Ricardo da S. Torres²

¹ CPqD Foundation

Rod. Campinas - Mogi-Mirim km 118.5 Campinas, SP, Brazil

² Institute of Computing, University of Campinas

Av. Albert Einstein, 1251 Campinas, SP, Brazil

fbellette@gmail.com, nadiapk@ic.unicamp.br, rtorres@ic.unicamp.br

Abstract. Biodiversity studies are often based on the use of data associated with field observations. These data are usually associated with a geographic location. Most of existing biodiversity information systems provides support for storing and querying geographic data. Annotation services, in general, are not supported. This article presents an annotation web service to correlate biodiversity data with geographic information. We use superimposed information concepts for constructing a Web service for annotating vector geographic data. The Web service specification includes the definition of a generic API for handling annotations and the definition of a data model for storing them. The solution was validated through the implementation of a prototype for the biodiversity domain considering a potential usage scenario.

Categories and Subject Descriptors: H.2.8 Information Systems [**Database Management**]: Database applications

Keywords: Annotation, Geographic Data, Web Service, Biodiversity

1. INTRODUCTION

The term biodiversity – or biological diversity – describes the richness and variety of biological organisms in a given habitat. Some known issues in biodiversity are irreversible loss of species, loss of environmental services (production of oxygen by plants, the hydrological balance, soil fertility, and climate balance), and biopiracy.

Computer Science can be a great allied of biologists, providing them with tools to analyze and report findings on species and their behaviors. Some challenges for the Biodiversity Information Systems (BISs) are [da S. Torres et al. 2006]: (i) handle large volumes of information, (ii) integrate information from different sources and formats (heterogeneity), (iii) manipulate data and images, (iv) manipulate geospatial information reference.

In this article we address the fourth challenge. Biodiversity studies are often based on the use of data associated with field observations, later matched with geographic locations. Most of existing biodiversity information systems provide support for storing and querying geographic data. Annotation services, in general, are not supported.

Annotation has been recognized as one of the most important services in digital library systems to foster the cooperation among users and the integration of heterogeneous information resources [Agosti and Ferro 2008]. In this article, we describe a new geographic data annotation Web service, named GeoNote, that can be easily integrated with other applications. The specification and implementation of the proposed Web service relies on two main contributions: (a) the proposal of a data model based on superimposed information [Maier and Delcambre 1999] to manage geographic annotations; and (b) the definition of a generic API to manipulate annotations.

We would like to thank CAPES, FAPESP, and CNPq (in particular, the BioCORE Project).

Copyright©2011 Permission to copy without fee all or part of the material printed in JIDM is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

The rest of this article is organized as follows. Section 2 presents the main concepts behind our solution. Section 3 describes the proposed geographic data annotation service. Section 4 describes the use of the proposed annotation service in a Biodiversity Information System. Section 5 briefly comments on related research. Section 6 presents conclusions comments on related research. Section 6 presents conclusions and summarizes ongoing and future work.

2. BACKGROUND

This section describes the main concepts associated with the proposed annotation service.

2.1 Superimposed Information

Superimposed information (SI) [Maier and Delcambre 1999] is the new information (or new interpretation) associated with the existing information or base (as for example, an annotation, term or keyword, a comment about a base document). In general, SI is not structured and does not have a pre-defined schema. Different types of SI can be associated with the same base information [Delcambre and Maier 1999]. SI can be used in several ways: to help locating an information, as an index or content tables; to explain, add, correct or deny an information; to relate and compare different information; to classify and organize information.

SI references the base information using a “mark”. A mark can be used to activate the referenced information (*excerpt*) in its original context (*context*). Context information is a set of information obtained using a mark. For example, the page number and font characteristics are part of the context information from a selection in a PDF file.

2.2 Geographic Data

The term spatial data denotes any type of data that describe phenomena associated with some spatial dimension. Geo-referenced data belong to a particular class of spatial data, describing events, objects, and phenomena on the globe associated with their location on the surface, at a certain moment or period of time [Câmara et al. 1996].

2.2.1 Data Models. Within the context of applications which deal with geographic information, the real world is often modeled under two complementary views: the field-based model and object-based model. The field-based model (usually implemented according to the structure called *raster*) deals with a continuous surface on which the geographic phenomena varies according to different distributions (like atmospheric pressure). Each layer corresponds to a different theme (like vegetation, soil) [Gomes Jr 2007].

The object-based model (usually implemented using a vector representation) represents the world as an area occupied by identifiable objects that have their own geometry and characteristics. In the object-based model, a geographic object is typically represented in vector format, that is, its geometry is described using points, lines, and polygons. Lines are formed by sequences of points and polygons (open or closed), by a sequence of lines [Gomes Jr 2007]. The current version of the proposed annotation service is focused on vector data.

2.2.2 Spatial Operators. Several spatial operators have been proposed to process geographic objects based on their relationships. These relationships can be divided into three categories: metric, topological, and directional [Güting 1994].

Examples of topological relationships include *equals*, *disjoint*, *intersects*, *touches*, *crosses*, *overlaps*, and *contains*. Figure 1 illustrates some of the possible topological relationships. Directional relationships include, for example, *above* and *north_of*. Metric relationships include *distance* and *area*. Spatial operators are used in our Web service to define which geographic objects (or parts of them) are being annotated.

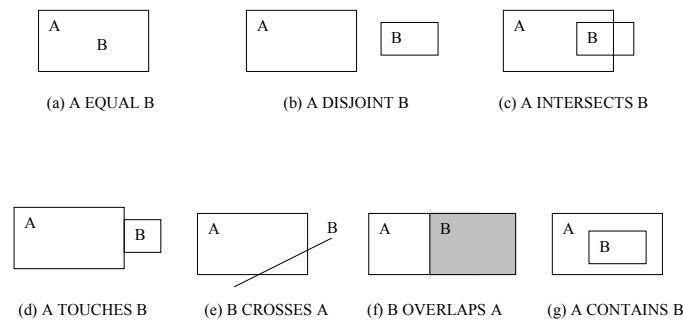


Fig. 1. Examples of topological relationships: *equals*, *disjoint*, *intersects*, *touches*, *crosses*, *overlaps*, and *contains*.

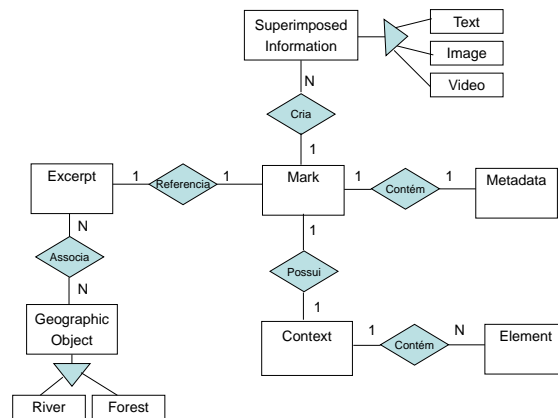


Fig. 2. Proposed entity-relationship diagram to manage annotations associated with geographic objects.

3. GEONOTE: A WEB SERVICE FOR ANNOTATING GEOGRAPHIC DATA

This section describes the proposed Web service for annotating geographic data. Section 3.1 presents how we have extended superimposed information concepts to handle geographic data annotations. Section 3.2 presents the main operations supported by GeoNote.

3.1 Data Model

Figure 2 shows the entity-relationship diagram that represents the data model used for constructing the annotation Web service. The entities are represented by rectangles. Relationships are represented by diamonds. Lines represent the connections between the entities and relationships. Triangles indicate generalizations. The cardinality of the relationship is showed next to the diamonds.

The data model is based on SI concepts. The entities *Excerpt*, *Mark*, *SuperimposedInformation*, *Context*, and *Element* are inspired by the SPARCE model [Murthy et al. 2004; Murthy et al. 2004].

When an annotation is inserted, a mark is created. A mark is used to link the superimposed information with the base information [Archer et al. 2008]. Different types of marks can be created when dealing with maps. Figure 3 illustrates four examples of marks associated with geographic data. The mark showed in Figure 3(a) refers to a rectangular region that includes several geographic objects represented as lines and points. In Figure 3(b), the mark includes parts of Brazilian states (i.e., parts of polygons). Figure 3(c) illustrates the definition of a mark associated with the result of applying a spatial operator (*intersection*) on two regions

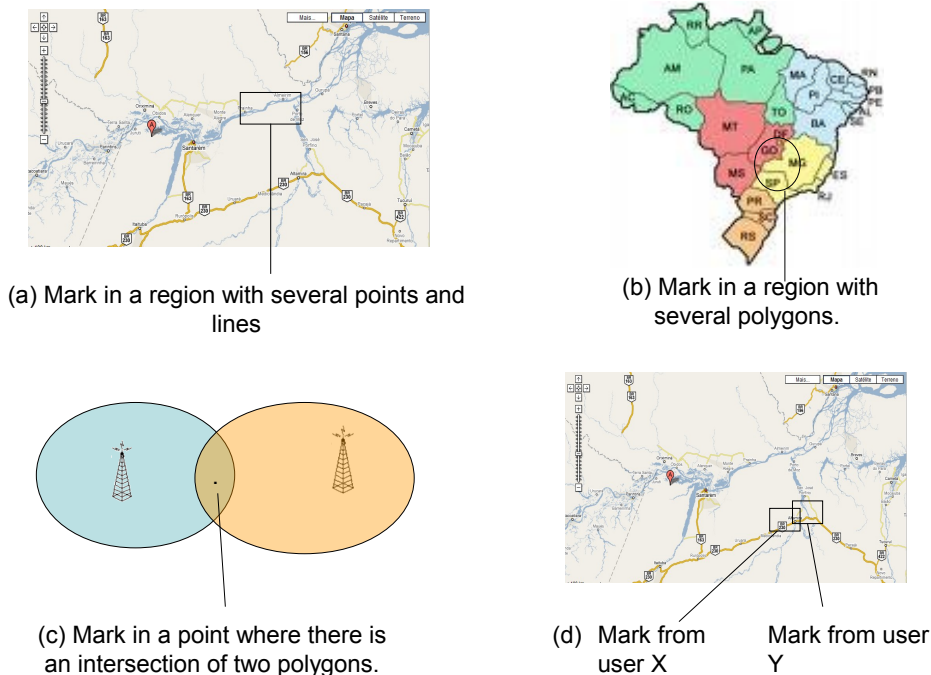


Fig. 3. Examples of marks.

defined by the coverage of two antennas (i.e., intersection of two polygons). Finally, Figure 3(d) illustrates how to define a new mark associated with the intersection of the regions defined by two other marks (in the case of the example, marks defined by users X and Y).

An *Excerpt* represents the content of a marked region. In the case of vector data, this content may be a point, a line (or part of a line), a polygon (or part of a polygon), or sets of objects with different geometries. Figures 4, 5, 6, and 7 illustrate how excerpts are created, given defined marks.

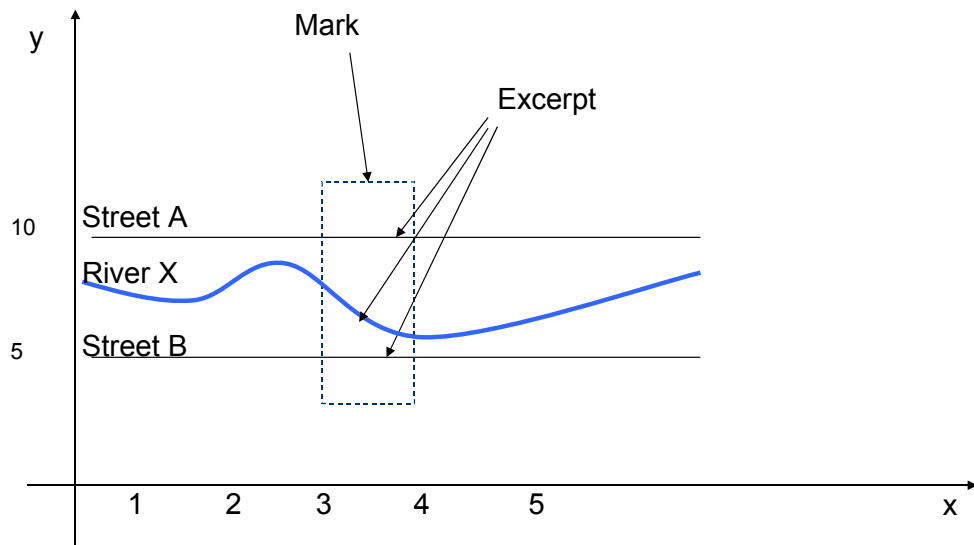
The *Context* entity defines a set of *elements* that provide contextual information associated with a mark. In the case of textual base document, context data may include, for example, information about font type, font size, or page number. For geographic data, context may involve information related to the geographic coordinate system or visual properties (e.g., color, size, length) used for visualization purposes.

Metadata attributes relies on *Dublin Core* model elements¹. Metadata includes information about the annotation creator, language, annotation date, and rights held in and over the geographic data being annotated.

Superimposed Information refers to a new information associated with the information base (vector geographic data). In the current version of our annotation service, SI refers to textual annotations.

The spatial components of the data model (not shown in Figure 2 for clarity reasons) follow the Open Geospatial Consortium (OGC) model [OGC 2010]. OGC defines the *Geometry* as the most generic entity. The geometry is associated with a Spatial Reference System, which specifies the coordinate system. The basic representations of geometry are *point*, *curve*, *surface*, and *collection*. A *GeometryCollection* can be used to represent a collection of distinct geometries. *MultiPoint*, *MultiLineString*, *MultiPolygon*, *MultiCurve*, and *MultiSurface* are homogeneous geometry collections of *Point*, *LineString*, *Polygon*, *Curve*, and *Surface*, respectively. A *LineString* represents a sequence of connected points. A *Point* represents a simple location in the coordinate space. A *Curve* is a generalization of *LineString*. A *Line* is a specialization of *LineString*

¹<http://dublincore.org>



IdGeographic Object	Name	Type	The_Geometry
1	A	Street	0,10; 10,10
2	B	Street	0,5; 10,5
3	X	River	0,7; 10,7

IdGeographic Object_has_Excerpt	IdExcerpt	Id Geographic Object	The_Geometry
101	100	1	3,10; 4,10
102	100	2	3,5; 4,5
103	100	3	3,7; 4,6

Fig. 4. Example of the execution of the operation *boolean InsertAnnotation (Text Annotation, Geometry Region, Metadata Metadata, Context Context)*.

which has only two points. A *LinearRing* represents a *LineString* which is closed and complete. A *Polygon* is a closed plane figure, defined by one exterior boundary and zero or more interior boundaries. A *Surface* is a generalization of a *Polygon*. We assume that application geographic objects (e.g., river and street) have a *Geometry*.

3.2 Web Service API

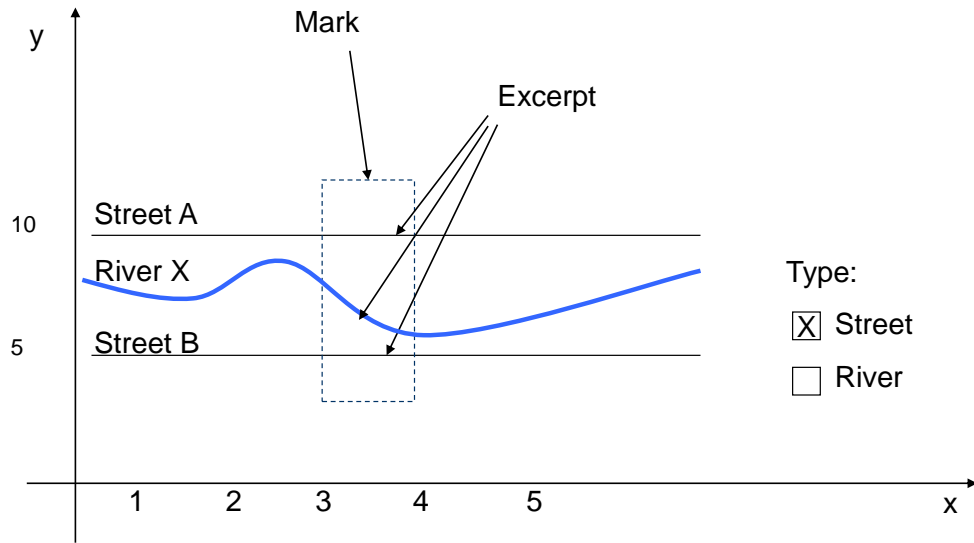
With the data model presented in the previous section, we now present the interfaces for the insertion, update, deletion, and search operations.

A generic interface was defined to manage annotations:

< return > operation (**Text** Annotation, **Query** Query, **Metadata** Metadata, **Context** Context)

where *operation* can assume the insert (*InsertAnnotation*), update (*UpdateAnnotation*), delete (*DeleteAnnotation*), or query (*QueryAnnotation*) operations. The *Annotation* parameter refers to the annotation content (in the case of the current version of our Web service, text annotation). The *Query* parameter defines a query whose results include the set of geographic objects to which the target operation will be related. The *Metadata* parameter refers to attributes associated with the mark metadata. The *Context* parameter includes elements of the mark context.

Note that since several parameters are managed, the interface was later specialized for geo-queries (regions, excerpts, etc.), non-spatial queries (metadata, annotation, etc.) and their combination. One remarkable aspect



IdGeographic Object	Name	Type	The_Geometry
1	A	Street	0,10; 10,10
2	B	Street	0,5; 10,5

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_Geometry
101	100	1	3,10; 4,10
102	100	2	3,5; 4,5

Fig. 5. Example of the execution of the operation *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **List** {ObjectType Type}, **Metadata** Metadata, **Context** Context).

of the proposed API relies on the support of operations which are defined over geographic objects that are result of queries.

3.2.1 *Annotation Insertion*. This section presents the APIs of functions used to insert annotations. The interfaces are derived from the generic interface presented in Section 3.2.

—**boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **Metadata** Metadata, **Context** Context)

This operation inserts an annotation associated with a set of geographic objects within a given region. Figure 4 illustrates the insertion of an annotation associated with the region defined by the rectangular region (mark). As it can be observed, this annotation is linked to part of Street A, part of River X, and part of Street B (*Excerpt*).

Table *IdGeographicObject* contains records related to geographic objects (e.g., Street A, Street B, and River X). Table *IdGeographicObject_has_Excerpt* contains records related to the relationship of geographic objects and excerpts, after inserting a mark.

—**boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **List** {ObjectType Type}, **Metadata** Metadata, **Context** Context)

This operation inserts an annotation associated with a set of geographic objects that can be found in a given region and that are of a given type. Figure 5 illustrates this insertion operation. As it can be observed, in this case, the annotation is linked to geographic objects of type Street (*Excerpt*).

—**boolean** *InsertAnnotation* (**Text** Annotation, **Geometry** Region, **Query** Query, **Metadata** Metadata, **Context** Context)

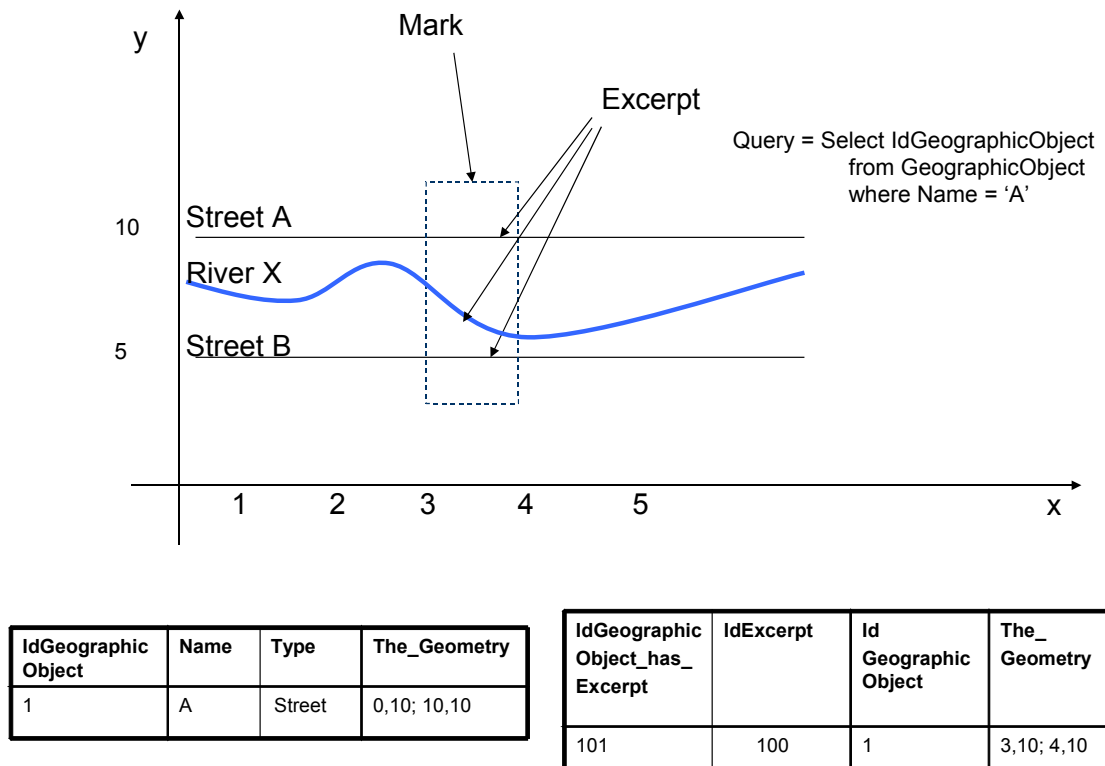


Fig. 6. Example of the execution of the operation *boolean InsertAnnotation* (*Text Annotation*, *Geometry Region*, *Query Query*, *Metadata Metadata*, *Context Context*).

This operation inserts an annotation associated with a set of geographic objects defined as the result of the query specified in the parameter *Query*. These geographic objects must be found in the region defined by parameter *Region*.

Figure 6 illustrates the insertion of an annotation associated with the region defined by the dashed blue rectangle (mark). As it can be observed, in this case, the annotation is linked to only parts of Street A (*Excerpt*), which is the result of the query specified by the parameter *Query*.

—**boolean InsertAnnotation** (**Text Annotation**, **GeoQuery GeoQuery**, **List {ObjectType Type}**, **Metadata Metadata**, **Context Context**)

This operation inserts an annotation associated with the geographic objects of a given type that satisfy the query defined in parameter *GeoQuery*.

—**boolean InsertAnnotation** (**Text Annotation**, **GeoQuery GeoQuery**, **Metadata Metadata**, **Context Context**)

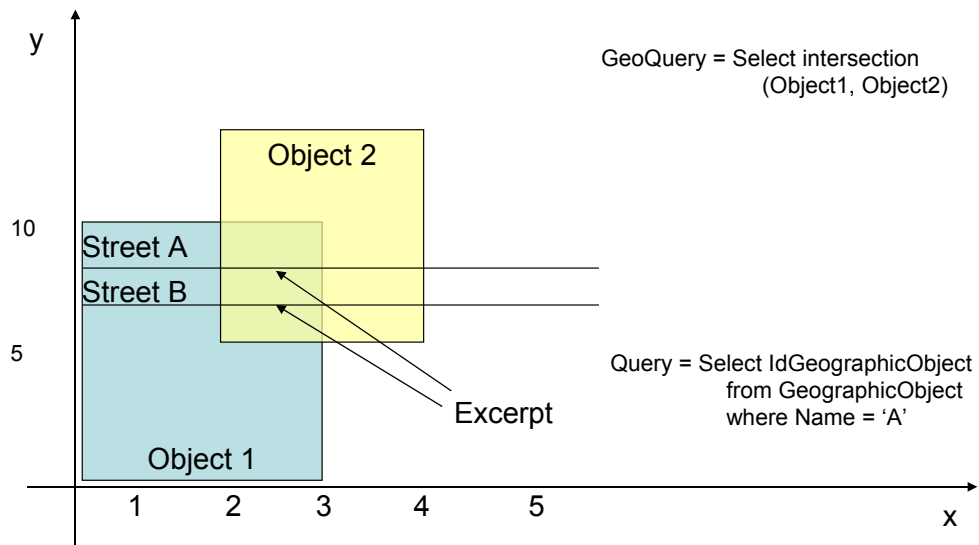
This operation inserts an annotation associated with geographic objects defined as the result of the geographic query specified in parameter *GeoQuery*.

Figure 7 illustrates the insertion of an annotation associated with the region defined by the rectangular region (mark). As it can be observed, in this case, the annotation is linked to only parts of Streets A and B (*Excerpt*), which are the result of the query specified by parameter *GeoQuery*. Note that this *GeoQuery* returns the region defined by the intersection of the geometries of Object 1 and Object 2.

3.2.2 *Annotation Deletion*. This section presents the APIs of functions used to delete annotations.

—**boolean DeleteAnnotation** (**List {MetadataField Field}**, **Operator Operator**, **MetadataValue Value**, **Connector Connector**)

This operation removes the metadata annotations given the parameters defined.



IdGeographic Object	Name	Type	The_Geometry
1	A	Street	0,9; 10,9
2	B	Street	0,8; 10,8

IdGeographic Object_has_ Excerpt	IdExcerpt	Id Geographic Object	The_Geometry
101	100	1	2,9; 3,9
102	100	2	2,8; 3,8

Fig. 7. Example of the execution of the operation *boolean InsertAnnotation (Text Annotation, GeoQuery GeoQuery, List {ObjectType Type}, Metadata Metadata, Context Context)*.

—**boolean DeleteAnnotation (List {String Keyword})**

This operation removes the annotations that contain the given keywords.

—**boolean DeleteAnnotation (Geometry Region)**

This operation removes the annotations found in a given region.

—**boolean DeleteAnnotation (Query Query)**

This operation removes the annotations associated with geographic objects that satisfy a given query.

—**boolean DeleteAnnotation (GeoQuery GeoQuery)**

This operation removes the annotations from geographic objects that satisfy a given geographical query.

3.2.3 *Annotation Update*. This section presents the APIs of functions used to update annotations. The interfaces are derived from the generic interface presented in Section 3.2.

—**boolean UpdateAnnotation (Text NewAnnotation, List {MetadataField Field, Operator Operator, MetadataValue Value, Conector Conector})**

This operation updates annotations whose metadata coincide with the parameters provided.

—**boolean UpdateAnnotation (Text NewAnnotation, List {String Keyword})**

This operation updates annotations which contain the given keywords.

—**boolean UpdateAnnotation (Text NewAnnotation, Geometry Region)**

This operation updates annotations found in a given a region.

- boolean** *UpdateAnnotation* (**Text** NewAnnotation, **Query** Query)
This operation updates annotations associated with geographic objects that satisfy the query defined in *Query*.
- boolean** *UpdateAnnotation* (**Text** NewAnnotation, **GeoQuery** GeoQuery)
This operation updates annotations associated with geographic objects that satisfy the geographic query defined in *GeoQuery*.
- boolean** *UpdateAnnotation* (**Text** OldAnnotation, **Text** NewAnnotation)
This operation replaces an annotation by a new one.
- boolean** *UpdateMark* (**MarkGeometry** NewMark, **MarkGeometry** OldMark)
This operation updates the coordinates of an annotation mark.

3.2.4 *Query*. This section describes the interfaces for querying annotations, considering a list of keywords, an object, a region, or results from a query.

- List** {**Text**, **Metadata**} *QueryAnnotation* (**List** {**MetadataField** Field, **Operator** Operator, **MetadataValue** Value, **Connector** Connector})
This operation returns the annotations (and associated metadata), which satisfy the given parameters.
- List** {**Text**, **Metadata**} *QueryAnnotation* (**List** {**String** Keywords})
This operation returns the annotations which contain a given list of keywords.
- List** {**Text**, **Metadata**} *QueryAnnotation* (**String** ObjectName)
This operation returns the annotations which are associated with a given geographic object.
- List** {**Text**, **Metadata**} *QueryAnnotation* (**Geometry** Region)
This operation returns the annotations found in a given region.
- List** {**Text**, **Metadata**} *QueryAnnotation* (**Query** Query)
This operation returns the annotations associated with objects defined by the results of the query defined in *Query*.
- List** {**Text**, **Metadata**} *QueryAnnotation* (**GeoQuery** GeoQuery)
This operation returns the annotations associated with geographic objects defined by the results of the geographic query *GeoQuery*.
- MarkGeometry** *QueryMark* (**Text** Annotation)
This operation returns the geometry of a mark whose annotation includes terms found in parameter *Annotation*. This operation can be used, for example, to obtain the geometries of the marks defined by users X and Y, as illustrated in Figure 3(d).

4. CASE STUDY: BIODIVERSITY INFORMATION SYSTEM

This section presents a case study that concerns with the use of GeoNote to handle biological data related to biodiversity studies. In our case study, we consider the use of the proposed annotation service to associate field observations on species with geographic information.

4.1 Data Source

We have implemented a Biodiversity Information System for managing *ocurrence records*, which contain observations of butterflies collected during field trips to the Municipal Reserve of Santa Genebra, a remnant forest fragment in Campinas, Brazil.

The biodiversity database model proposed in [Malaverri et al. 2009] was used for storing field observations. This data model considers the following elements: species taxonomic information, description of collection methodology and data on how, where, when, and by whom butterflies were collected.

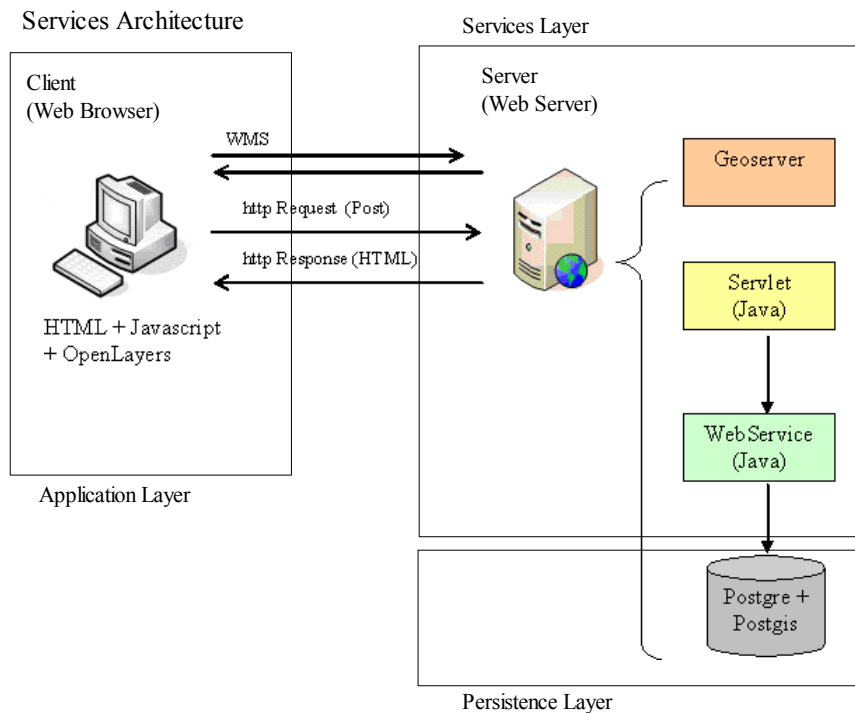


Fig. 8. Architecture of the implemented Biodiversity Information System.

In our prototype, we consider the data provided by the study described in [Gomes Jr. 2003]. The main objective of that study is to describe quantitatively the community of frugivorous butterflies with an emphasis on populations of *Anaea ryphea* (Cramer) (Nymphalidae: Charanidae). Samples were held for 26 months. The data acquisition procedure considered the distribution of 12 traps in three areas with different levels of disturbance: 4 at the reserve boundaries, 4 in the forest interior, and 4 near a central path (abandoned road), which crosses the whole reserve. The objective of this study is to evaluate butterfly behaviour features, given different luminosity, temperature, and wind conditions.

In this prototype, only text annotations were considered. Initially, the database was loaded with data obtained from digital maps of Brazil in Shapefile format (available at <http://www.gismaps.com.br/>). The natural and environment resources chosen were: Geography, Preservation Areas, and National and State Parks.

4.2 Architecture of the Biodiversity Information System

Figure 8 shows the architecture of the implemented Biodiversity Information System. This architecture is composed of three layers: Application, Services, and Persistence.

The Application layer refers to a client software that will access the service layer. It is responsible for the interaction with the user. Usually, it can be a Web interface or a component of another system. In the case of our prototype, the client application layer was composed by HTML pages and Javascript running in a Web browser. The rendering of the maps was performed on the server. The OpenLayers framework was used for map presentation. OpenLayers is an open source JavaScript library used to display spatial data in web pages. The OpenLayers API allows several tools to interact with the visual maps or performing actions on certain events. The HTML page invokes a servlet passing the desired operation and the required parameters.

The Service layer provides the APIs, which define how features can be invoked by other applications. This

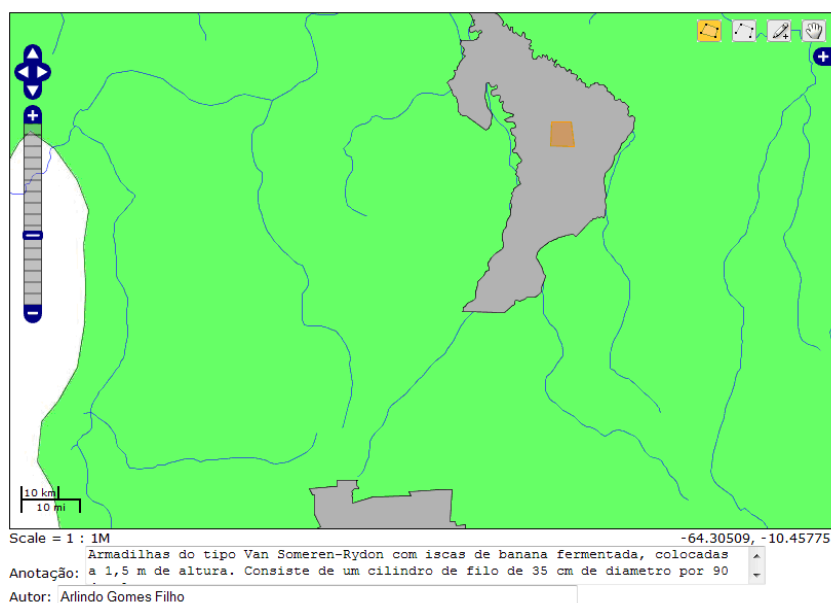


Fig. 9. Example of annotation of a region within an ecological reserve.

layer is composed by two components: a map server that is used to compose and render a map to be published on the web; and the proposed geographic annotation service that is used to insert annotations associated with geographic data. Geoserver² was used to implement the map server. GeoServer is compliant with the OGC open standards such as Web Map Service (WMS), Web Coverage Service (WCS), and Web Feature Service (WFS-T).

The Persistence layer includes database facilities to manage data about geographic objects, annotations, and marks. This layer uses PostGreSQL³ database system and PostGIS⁴. PostGIS can be seen as an OpenGIS conformant extension to the PostgreSQL, which allows geographic information systems objects to be stored into the database. Furthermore, PostGIS provides mechanisms for processing spatial operators.

4.3 Usage Scenario

Figure 9 illustrates the creation of an annotation. In the example, a polygon was drawn within an ecological reserve (orange polygon). Next, the user inserts an annotation indicating that the defined polygon contains special traps for capturing butterflies. In this case, the excerpt is a subregion of the ecological reserve.

A biologist may also insert annotations on traps located at the boundaries of the reserve. Figure 10 illustrates this kind of annotation. In this example, the polygon was drawn around the border and is partly contained in the area of the ecological reserve. The excerpt is a subregion of the ecological reserve which *intersects* the defined mark region. The intersection operation takes advantage of the *InsertAnnotation* operation that uses a geographic query.

Figure 11 illustrates how a biologist can query annotations. In this case, a polygon (region of interest – Municipal Reserve of Santa Genebra) is defined and annotations associated with geographic objects found within the polygon are returned (bottom part of the figure).

Figure 12 illustrates how a biologist can query annotations whose marks are contained within an ecological

²<http://geoserver.org/display/geos/welcome>

³<http://www.postgresql.org>

⁴<http://postgis.refractory.net>

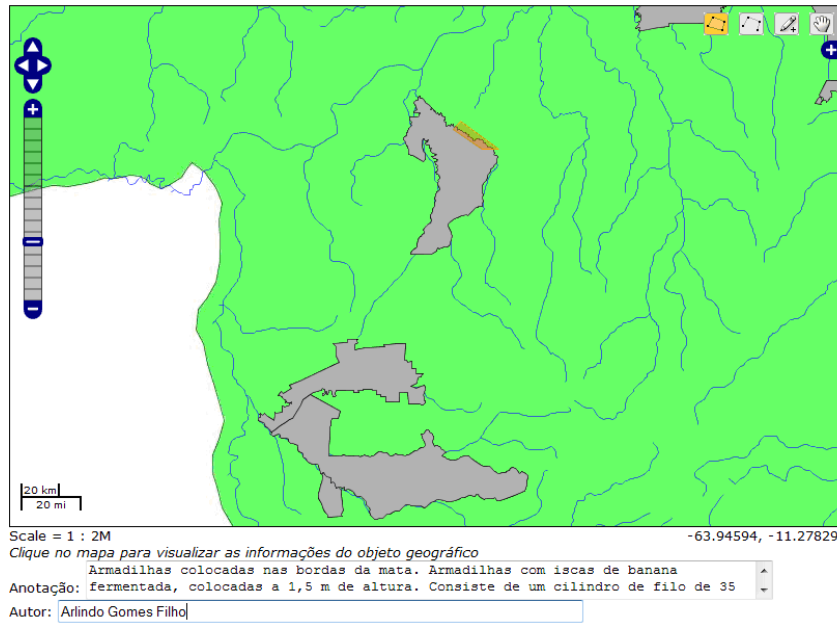


Fig. 10. Example of annotation around the border of an ecological reserve.

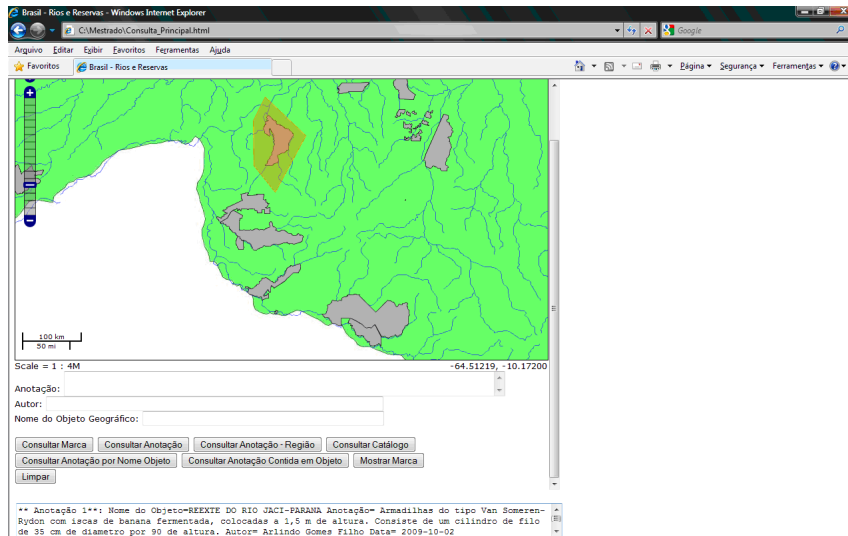


Fig. 11. Example of a query for annotations found within a region of interest.

reserve. The input parameter of the query is the name of the geographical object, which was provided in the corresponding field of the screen. In the example, the object was the geographical reserve REEXTE near the JACI-PARANA river. Only a part of the name (*jaci*) was informed as query parameter.

Figure 13 shows part of a WSDL specification related to the annotation query operation that is based on the name of a geographic object. The *input* and *output* sections define which parameters are sent and returned by SOAP messages when this operation is executed.

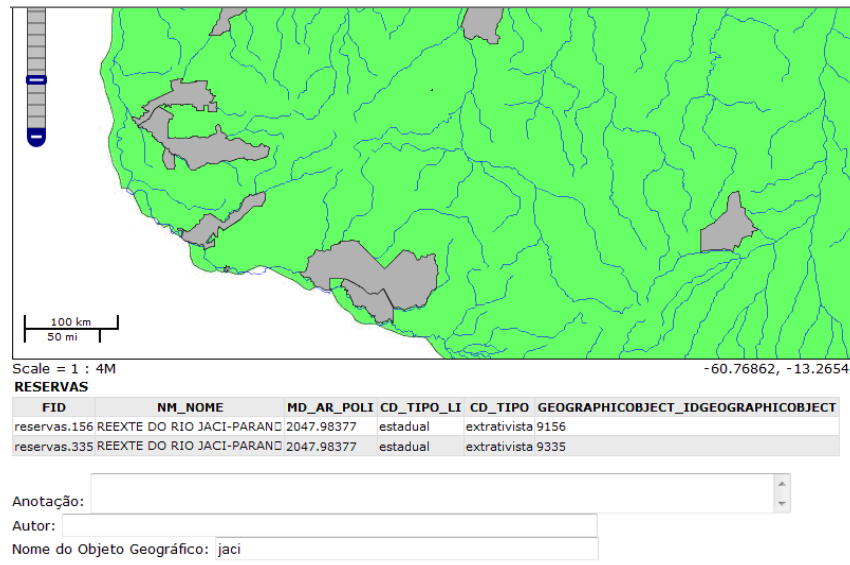


Fig. 12. Example of a query for annotations given a textual term.

```

<definitions xmlns:soap = "http://schemas.xmlsoap.org/wsdl/soap/" xmlns:tns = "teste" xmlns:xsd =
"http://www.w3.org/2001/XMLSchema" xmlns = "http://schemas.xmlsoap.org/wsdl/" targetNamespace =
"teste" name = "Annotation_QueryService">
  <types>
    <xsd:schema>
      <xsd:import namespace = "teste" schemaLocation =
"http://localhost:8080/queryByObjName?xsd-1"/>
    </xsd:schema>
  </types>
  <message name = "queryByObjName">
    <part name = "parameters" element = "tns:queryByObjName"/>
  </message>
  <message name = "queryByObjNameResponse">
    <part name = "parameters" element = "tns:queryByObjNameResponse"/>
  </message>
  <portType name = "Annotation_Query">
    <operation name = "queryByObjName">
      <input message = "tns:queryByObjName"/>
      <output message = "tns:queryByObjNameResponse"/>
    </operation>
  </portType>
  <binding name = "Annotation_QueryPortBinding" type = "tns:Annotation_Query">
    <soap:binding transport = "http://schemas.xmlsoap.org/soap/http" style = "document"/>
    <operation name = "queryByObjName">
      <soap:operation soapAction = ""/>
      <input>
        <soap:body use = "literal"/>
      </input>
      <output>
        <soap:body use = "literal"/>
      </output>
    </operation>
  </binding>
  <service name = "Annotation_QueryService">
    <port name = "Annotation_QueryPort" binding = "tns:Annotation_QueryPortBinding">
      <soap:address location = "http://localhost:8080/queryByObjName"/>
    </port>
  </service>
</definitions>

```

Fig. 13. Example of WSDL specification of the annotation query operation.

5. RELATED WORK

The research described in this article differs from related research in the sense that it takes advantage of superimposed information concepts to provide mechanisms for geographic data annotation in biodiversity applications. Furthermore, the use of a Web service implementation ensures appropriate encapsulation of data and procedures, which allows reuse of the developed component in other DL initiatives involving geographic data annotation.

Superimposed information has been used on several applications, which are named *Superimposed applications (SAs)* [Maier and Delcambre 1999; Murthy et al. 2004; Murthy et al. 2006]. They allow the manipulation and creation of superimposed information, typically to highlight, annotate, draw, select, organize, connect, or reuse the information [Murthy et al. 2006]. Examples include RIDPad [Murthy et al. 2004], Schematics Browser [Murthy et al. 2004], SIMPEL [Murthy et al. 2006], SLIMPad [Delcambre et al. 2001], and Mash-omatic [Murthy et al. 2006]. None of these SAs, however, support interactions with geographic data.

There are some other DL initiatives for the biodiversity domain. One example concerns floristic digital libraries (FDLs) [Sanchez et al. 1999]. FDLs use an agent-based infrastructure to manage botanical information about taxonomic keys, distribution maps, illustrations, and treatments (morphological descriptions). Another example is the Taiwanese digital museum of butterflies [Zhu et al. 2000]. This digital library offers a set of XML-based modules to manage data on butterflies. Geographic data annotation services, however, are not supported in these systems.

More recent DL-oriented initiatives include [Yu et al. 2006; Barros et al. 2008; Murthy et al. 2009; da S. Torres et al. 2006]. The *Superimposed Image Description and Retrieval Tool* (SuperIDR) [Murthy et al. 2009; Kozievitch et al. 2010; Murthy et al. 2011] is an extension of SIERRA [Murthy et al. 2006], designed to work with Tablet PC. With SuperIDR, users can navigate through the collection by species or taxonomic organization. Users can perform queries using part of an image or a textual description and recover the original content. Both systems work with annotation, but they still do not support geographic annotation. One first step on that direction is given in [Kozievitch et al. 2010].

Ecopod [Yu et al. 2006] aimed to replace the paper field guides with a mobile computing platform for identifying plants and animals in the field. In this tool, geographic information is used as an evidence in the species identification process. For example, some species are observed more frequently in specific regions. Annotation facilities, however, are not supported. Barros et al. [Barros et al. 2008] have proposed a georeferenced digital library, named BDIG-PELD that integrates network of ecological sites using OAI-PMH protocol and the ODL (Open Digital Library) framework. Even though georeferencing facilities are used to determine species location, annotation services are not provided. Torres et al. [da S. Torres et al. 2006] have proposed an architecture to combine research on image processing, databases, and digital libraries to provide biodiversity researchers with a BIS that integrates queries involving both image content and textual data. This architecture does not support services that handle geographic data.

There are also examples of applications which work with geographic data annotation [Macário et al. 2009]. Besides having several standards for services and implementations, these applications are still lacking of support for many features, like annotation of objects that are the result of a query.

6. CONCLUSIONS

We have presented the specification and implementation of a Web service for managing geographic data in Biodiversity Information Systems. The service manages information in text format, referred to as annotations. Two main contributions are provided: the definition of a data model based on Superimposed Information concepts to deal with geographic data annotations and the specification and implementation of a set of operations to manage annotations. One remarkable aspect of the proposed API relies on the support of operations (e.g., insertion, deletion, query, update, and deletion) which are defined over geographic objects that are result of queries. In fact, spatial operators can be used to determine which geographic objects (or parts of them) should be annotated.

A generic interface for handling annotations was defined, generating other specific interfaces to facilitate the use of the proposed annotation service. To validate our ideas, we developed a Biodiversity Information System that takes advantage of the proposed annotation service to correlate ecological data with geographic information.

Ongoing and future work includes the extension of the proposed API to handle not only vector data, but also

raster. We also plan to incorporate the implemented annotation service into the SuperIDR tool⁵ [Murthy et al. 2009] and into BioCore infrastructure⁶. Another future work consists of applying the proposed service in other domains (e.g., Agriculture). Finally, we plan to conduct user evaluation experiments to assess the effectiveness and efficiency of the annotation service.

REFERENCES

- AGOSTI, M. AND FERRO, N. Annotations: A way to interoperability in DL. In *Proceedings of the European Conference on Digital Library*. Aarhus, Denmark, pp. 291–295, 2008.
- ARCHER, D. W., DELCAMBRE, L., CORUBOLO, F., CASSEL, L., PRICE, S., MURTHY, U., MAIER, D., FOX, E. A., MURTHY, S., MCCALL, J., KUCHIBHOTLA, K., AND SURYAVANSHI, R. Superimposed Information Architecture for Digital Libraries. In *Proceedings of the European Conference on Digital Library*. Aarhus, Denmark, pp. 88–99, 2008.
- BARROS, E. G., LAENDER, A. H. F., GONÇALVES, M. A., AND BARBOSA, F. A. R. A digital library environment for integrating, disseminating and exploring ecological data. *Ecological Informatics* 3 (4-5): 295–308, 2008.
- CÂMARA, G., CASANOVA, M. A., MAGALHÃES, G. C., AND MEDEIROS, C. M. B. *Anatomia de Sistemas de Informação Geográfica*. 10a Escola de Computação, Instituto de Computação, Unicamp, Campinas, 1996.
- DA S. TORRES, R., MEDEIROS, C. M. B., CALVES, M. A. G., AND FOX, E. A. A Digital Library Framework for Biodiversity Information System. *International Journal on Digital Libraries* 6 (1): 3–17, 2006.
- DELCAMBRE, L. M. L. AND MAIER, D. Models for superimposed information. In *Proceedings of the Workshops on Evolution and Change in Data Management, Reverse Engineering in Information Systems, and the World Wide Web and Conceptual Modeling*. Paris, France, pp. 264–280, 1999.
- DELCAMBRE, L. M. L., MAIER, D., BOWERS, S., WEAVER, M., DENG, L., GORMAN, P., ASH, J., LAVELLE, M., AND LYMAN, J. Bundles in captivity: An application of superimposed information. In *Proceedings of the International Conference on Data Engineering*. Heidelberg, Germany, pp. 111–120, 2001.
- GOMES JR., A. *A Comunidade de Borboletas Frugívoras da Reserva de Santa Genebra, Campinas, SP, Com Ênfase na Flutuação Populacional de Anaea Ryphea (Cramer) (Nymphalidae: Charanidae) e Sua Relação com as Plantas Hospedeiras*. M.S. thesis, Unicamp, 2003.
- GOMES JR, L. C. *Uma arquitetura para consultas a repositórios de biodiversidade na Web*. M.S. thesis, Unicamp, 2007.
- GÜTING, R. H. An introduction to spatial database systems. *The VLDB Journal* 3 (4): 357–399, 1994.
- KOZIEVITCH, N. P., DA S. TORRES, R., ANDRADE, F., MURTHY, U., FOX, E. A., AND HALLERMAN, E. A teaching tool for parasitology: Enhancing learning with annotation and image retrieval. In *Proceedings of the European Conference on Digital Library*. Glasgow, UK, pp. 466–469, 2010.
- KOZIEVITCH, N. P., MURTHY, U., FOX, E. A., HALLERMAN, E., AND DA S. TORRES, R. A geographic annotation service in SuperIDR. In *VT-GIS 2010 Symposium*. Blacksburg, VA, 2010.
- MACÁRIO, C. G. N., DE SOUSA, S. R., AND MEDEIROS, C. B. Annotating geospatial data based on its semantics. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Seattle, Washington, pp. 81–90, 2009.
- MAIER, D. AND DELCAMBRE, L. M. L. Superimposed information for the internet. In *Proceedings of the ACM SIGMOD Workshop on The Web and Databases*. Philadelphia, Pennsylvania, USA, pp. 1–9, 1999.
- MALAVERRI, J. G., VILAR, B. S. C. M., AND MEDEIROS, C. B. A tool based on web services to query biodiversity information. In *Proceedings of the International Conference on Web Information Systems and Technologies*. Lisbon, Portugal, pp. 305–310, 2009.
- MURTHY, S., DELCAMBRE, L. M. L., AND MAIER, D. Explicitly representing superimposed information in a conceptual model. In *Proceedings of the International Conference on Conceptual Modeling*. Tucson, AZ, USA, pp. 126–139, 2006.
- MURTHY, S., DELCAMBRE, L. M. L., MAIER, D., AND BOWERS, S. Putting Integrated Information into Context: Superimposing Conceptual Models with SPARCE. In *Proceedings of the Asia-Pacific Conference on Conceptual Modelling*. Dunedin, New Zealand, pp. 71–80, 2004.
- MURTHY, S., MAIER, D., AND DELCAMBRE, L. M. L. Mash-o-matic. In *Proceedings of the ACM Symposium on Document Engineering*. Amsterdam, The Netherlands, pp. 205–214, 2006.
- MURTHY, S., MAIER, D., DELCAMBRE, L. M. L., AND BOWERS, S. Superimposed applications using SPARCE. In *Proceedings of the International Conference on Data Engineering*. Boston, MA, USA, pp. 861, 2004.
- MURTHY, U., AHUJA, K., MURTHY, S., AND FOX, E. A. Simpel: a superimposed multimedia presentation editor and player. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Chapel Hill, NC, USA, pp. 377, 2006.
- MURTHY, U., DA S. TORRES, R., AND FOX, E. A. SIERRA - a superimposed application for enhanced image description and retrieval. In *Proceedings of the European Conference on Digital Library*. Alicante, Spain, pp. 540–543, 2006.
- MURTHY, U., FOX, E. A., CHEN, Y., HALLERMAN, E., DA S. TORRES, R., RAMOS, E. J., AND FALCÃO, T. R. C. Superimposed image description and retrieval for fish species identification. In *Proceedings of the European Conference on Digital Library*. Corfu, Greece, pp. 285–296, 2009.

⁵<http://si.dlib.vt.edu/>

⁶<http://www.lis.ic.unicamp.br/projects/biocore/>

- MURTHY, U., LI, L. T., HALLERMAN, E., FOX, E. A., PÉREZ-QUIÑONES, M. A., DELCAMBRE, L. M. L., AND DA S. TORRES, R. Use of subimages in fish species identification: a qualitative study. In *Proceedings of the Joint International Conference on Digital Libraries*. Ottawa, ON, Canada, pp. 185–194, 2011.
- MURTHY, U., RICHARDSON, R., AND FOX, E. A. Enhancing Concept Mapping Tools Below And Above To Facilitate The Use Of Superimposed Information. In *Proceedings of the International Conference on Concept Mapping*. San Jose, Costa Rica, pp. 111–120, 2006.
- OGC. Opengis standards and related OGC documents, <http://www.opengeospatial.org/standards>, As of January 2010.
- SANCHEZ, J. A., FLORES, C. A., AND SCHNASE, J. L. Mutant: Agents as Guides for Multiple Taxonomies in the Floristic Digital Library. In *Proceedings of the ACM Conference on Digital Libraries*. Berkeley, California, USA, pp. 244–245, 1999.
- YU, Y., STAMBERGER, J. A., MANOHARAN, A., AND PAEPCKE, A. Ecopod: a mobile tool for community based biodiversity collection building. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Chapel Hill, NC, USA, pp. 244–253, 2006.
- ZHU, B., RAMSEY, M., AND CHEN, H. Creating a Large-Scale Content-Based Airphoto Image Digital Library. *IEEE Transactions on Image Processing* 9 (1): 163–167, 2000.