G-Portal: A Map-based Digital Library for Distributed Geospatial and Georeferenced Resources

Ee-Peng Lim* , Dion Hoe-Lian Goh, Zehua Liu
Wee-Keong Ng, Christopher Soo-Guan Khoo, Susan Ellen Higgins
Centre for Advanced Information Systems
School of Computer Engineering
Nanyang Technological University
Singapore, 639798
{aseplim, ashlgoh, aszhliu, awkng, assgkhoo, ashiggins}@ntu.edu.sg

ABSTRACT

As the World Wide Web evolves into an immense information network, it is tempting to build new digital library services and expand existing digital library services to make use of web content. In this paper, we present the design and implementation of G-Portal, a web portal that aims to provide digital library services over geospatial and georeferenced content found on the World Wide Web. G-Portal adopts a map-based user interface to visualize and manipulate the distributed geospatial and georeferenced content. Annotation capabilities are supported, allowing users to contribute geospatial and georeferenced objects as well as their associated metadata. The other features included in G-Portal's design are query support, content classification, and content maintenance. This paper will mainly focus on the architecture design, visualization and annotation capabilities of G-Portal.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: Systems Issues; J.2 [**Physical Sciences and Engineering**]: Earth and Atmospheric Sciences

General Terms

Design

Keywords

Digital Libraries. Geography. World Wide Web. Education.

*This work was conducted when Dr Lim visited the Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong.

JCDL'02, July 13-17, 2002, Portland, Oregon, USA.

1. INTRODUCTION

A web portal, or simply portal, has been commonly defined as a web site that provides well-organized information resources within a common domain. Traditionally, web resources within portals have been organized into various pre-defined categories and made available through links on web pages. For example, Yahoo! [15] is a general-purpose web portal that attempts to index a set of selected web resources into a large hierarchy of categories[1]. Web portals offer several known advantages to their users. Firstly, the well-categorized web resources allow users to quickly sift away large volumes of unwanted information and to focus on the relevant ones. Secondly, web portals often provide some search engines to query their indexed resources.

With an increasing amount of geospatial information accessible on the web, it is attractive to construct portals that organize these resources systematically and provide services over them. However, because these resources have spatial attributes (e.g. latitude and longitude), a more appropriate approach might be to organize them spatially on a map. Thus, web portals for such geospatial resources no longer just categorize and present them as textual data and links, but also as locations on a map. This is a natural mode of interaction as users of geospatial information typically locate information within specific regions of interest. Using a map-based approach, users will be able to zoom in quickly to their areas of interest and retrieve the desired resources based on location.

In this paper, we present **G-Portal**, a digital library project started at the Centre for Advanced Information Systems in Nanyang Technological University, Singapore. The aim of the project is to identify, classify and organize geospatial and georeferenced resources on the web and to provide digital library services (e.g. searching, visualization) for these types of resources.

Apart from extending the web portal concept to include geospatial and georeferenced information, the G-Portal project also seeks users' contribution and participation in two ways. Firstly, it provides the facilities for users to contribute meta-data description about geospatial and georeferenced resources. Secondly, it facilitates user collaboration and participation under different user-defined projects in which

[1]To be more precise, Yahoo! maintains its categories in a directed graph structure.

several groups of users can come together to contribute and share their web resources. They can also add their knowledge as annotations to the web resources if desired. By making annotations available among users within a group, knowledge sharing becomes possible over G-Portal.

1.1 A Use Case Example

To justify the design of G-Portal, we need to consider a wide range of use cases that suggest how such a web portal can be used by a user or a group of users. Each use case describes the steps required to accomplish a task using geospatial and georeferenced web resources. The development of such use cases for earth system education has been carried out in the DLESE (Digital Library for Earth System Education) research [13]. In the following scenario, we study the use case of an educator wanting to create a collection of resources about flora and fauna in Singapore for teaching purposes.

The educator first decides to get some graduate students' help in collecting the various species of flora and fauna in different locations found in Singapore. This information will be compiled and made available on the web. It is necessary for the graduate students to document the discovered flora or fauna species and their locations as they find them. The compiled information may involve images, text and other types of multimedia objects. This information, after some editing by the lecturer, will be made available to the undergraduate students and general public as reference materials. Note that now the lecturer would like the undergraduate students to read the materials and provide further comments as they learn more about the flora and fauna species found in Singapore.

The above use case is a fairly general one. It involves the description of what the general requirements are for using geospatial and georeferenced web resources. It also does not describe how the requirements can be met by a specific system design. Instead of trying to address the above requirements by a single system module, we have decided to have G-Portal provide a set of tools or modules that can be combined together in a flexible way to achieve the use case requirements.

In the above use case, information about users and user contributions in the form of web resources and annotations should be stored and managed. The format of the web resource description should be made common for the graduate students so that they are aware of the types of information to be contributed. One should also be able to group different web resources and annotations together under a common project. A classification scheme for the resources will thus be required for easy access, and these may be defined by the teacher.

The above suggests the need for a set of modules to deal with the subtasks. It turns out that modules can be designed to be generic enough to handle other use cases, such as collecting and organizing existing information about forests on the web, and finding and organizing web resources under a research topic e.g. global climate change.

1.2 Key Concepts

This section briefly introduces the key concepts of G-Portal. Some of the concepts will be discussed further in Section 3.

Resources are the basic elements of G-Portal. They are generic containers that store data. At its most fundamental level, a resource has an *identifier*, a *location*, a *schema pointer*, an *owner*, *access control information* and *contents and attributes* as defined in its schema. In G-Portal, resources are stored as XML documents.

Schemas are used to define the internal structure of resources. Every resource requires a schema. Resources with the same schema can be seen as instances of the schema, in which case, the schema plays the role of a Class in the Object-Oriented model. Schemas are not user-consumable data, but are meant to be used by G-Portal or other applications to interpret the contents of resources. Schemas are part of G-Portal but they will be exposed to external applications in the form of URLs for information exchange purposes.

In order to organize the various resources, users can define **Layers** that contain a collection of resources. These resources may or may not be of the same type (i.e. having the same schema). A resource may belong to different layers, if its access control information as specified by its owner allows it to be shared.

In addition, resources typically are not used in isolation but as part of a larger task. The concept of a **Project** is introduced to provide a higher level of abstraction and is used to define the collection of layers used for a particular task. Under one project, multiple layers can be defined. Each project will have one or more *core layers*, which serve as the spatial context for the whole project. A layer can only be owned and modified by one project but may be made available to other projects through replication.

The remaining sections of this paper present G-Portal in detail. Section 2 introduces the system architecture of G-Portal. Section 3 describes the management of geospatial and georeferenced resources in the system. Section 4 talks about how the system presents and visualizes the resources. Various related works are reviewed in Section 5. Section 6 concludes the paper and proposes ideas for future work.

2. G-PORTAL SYSTEM ARCHITECTURE

G-Portal is designed to support the creation, maintenance, visualization, and querying of geospatial and georeferenced web content. The overall system architecture is shown in Figure 1. G-Portal consists of eight major modules each responsible for different functions. They are the **user management**, **project management**, **schema management**, **visualization**, **resource presentation**, **query engine**, **web resource maintenance** and **resource classification** modules.

The user management module maintains a database of digital library users who can contribute resources to G-Portal. Each user is assigned a user profile that includes his/her username, password, and other information that may be used to personalize the content and services accessible by the user.

The project management module allows a user to define one or more projects each representing a logical organization of geospatial and georeferenced resources. There are essentially two kinds of resources in G-Portal. A resource may be a piece of information that can be located and uniquely identified on the web, or a piece of information contributed by a user. In the former case, a resource record is created with reference to the corresponding web information. In the

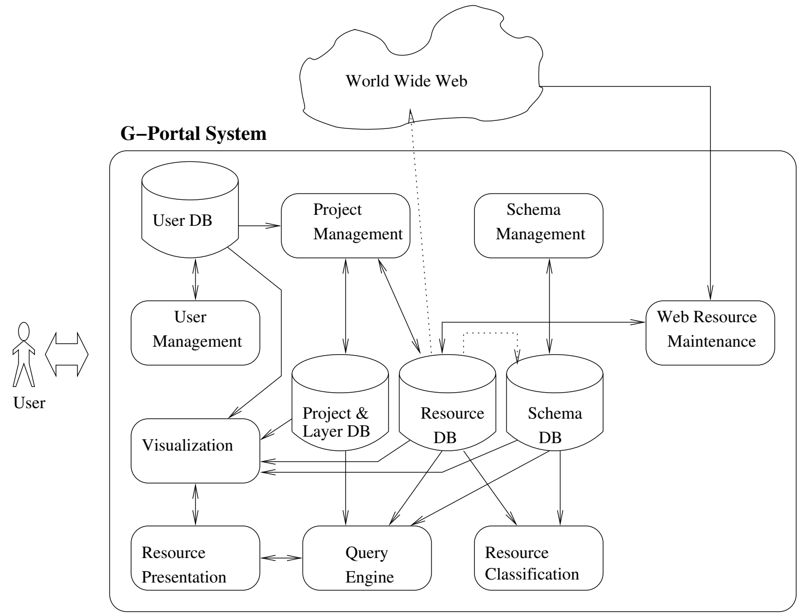

Figure 1: Architecture of G-Portal

latter, a resource record is also created but it will contain the contributed information as well. All resource records are stored in a **resource database**. Within each project are resources arranged in layers so that resources in the same layer are bundled and presented together. Note that resources can be shared across layers and projects but layers are not sharable among projects. The project and layer information are stored in the **project and layer database**. In Section 3, more details about resource management in G-Portal will be discussed.

To allow us to determine the semantics and structure of resources, each resource is defined by a schema. These schemas can be pre-defined or created as and when required. The schema management module is therefore responsible for the storage, retrieval, and querying on the schemas. Schema information is maintained in the **schema database.**

The visualization module is designed to provide a map-interface to the resources hosted by G-Portal. It consists of an interactive map that can be used to navigate and browse resources in a project. Nevertheless, such a visualization approach is only applicable to resources that carry spatial attributes. Consequently, all resources, spatial and non-spatial, will also be visualized in hierarchies of categories determined by the **resource classification module.** Depending on the schema involved, the resource presentation module will be able to display the contents of a selected resource. When the resource contains a URL reference to a web page, the resource presentation module can also directly invoke an external web browser to view the page. The other important features of the visualization module will be further elaborated in Section 4.

The query engine allows users to specify queries on projects, layers, resources, and resource schemas. As G-Portal adopts an XML database server to store the above information, the query language is XML-based and the query results will be formatted in XML as well. On the other hand, to support spatial queries over the resources, spatial indexes will be constructed on the spatial resources. For resources that are on remote database servers, the query engine will also provide meta-search services allowing users to query them directly.

The web resource maintenance module is mainly responsible for monitoring changes to the web information referenced by G-Portal's resources. When changes are detected, the resource owners will be notified so that they can either recapture the references or simply drop the resources. We anticipate this module will significantly increase the workload to the system. Hence, the maintenance feature should only be optional for resources having web references and will only be activated upon user specification.

3. RESOURCE MANAGEMENT

In this section, we describe how the various geospatial and georeferenced resources are managed in G-Portal.

3.1 Resources and Schemas

As mentioned in Section 1.2, a resource consists of several elements. The *identifier* is unique within the portal and is used to reference resources. The *location* of a resource associates the resource with either a location on a map or another resource. For the former, a coordinate system such as latitude and longitude may be used. The identifier of the associated resource will be used for the latter case. This occurs in situations where a resource is an annotation to another resource.

The *Schema pointer* specifies where to locate of the schema that defines the internal structure of the resource. The syntax of the pointer is based on a scheme similar to the Uniform Resource Locator (URL). Keeping the schema pointer within the resource itself allows viewers for the various types of resources to first read the resource schema, understand the syntactic structure of the resource and then retrieve and display the actual content.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema
    xmlns:xsd="http://www.w3.org/2000/10/XMLSchema"
    elementFormDefault="qualified">
  <xsd:include schemaLocation="Resource.xsd"/>
  <xsd:complexType name="FlowerType">
    <xsd:complexContent>
      <xsd:extension base="ResourceType">
        <xsd:sequence>
          <xsd:element name="Name" type="xsd:string"/>
          <xsd:element name="Type" type="xsd:string"/>
          <xsd:element name="Color" type="xsd:string"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>
  <xsd:element name="Resource">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="Flower" type="FlowerType"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Figure 2: A Resource Schema of Flowers

```
<?xml version="1.0" encoding="UTF-8"?> <Resource
    xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="Flower.xsd">
  <Flower ID="12BCk_2">
    ... ...
    <Name>Red Rose</Name>
    <Type>Rose</Type>
    <Color>Red</Color>
  </Flower>
</Resource>
```

Figure 3: An Instance of Flower Resource

The *owner* of a resource is the user who created the resource. Owners may specify the *access control information* of that resource (e.g. allow others to read but not modify the resource). Access control will be discussed later in Section 3.4.

The type of data that a resource stores is determined by the schema associated with the resource. The data might be as simple as a single text string or as complex as a tree structure of elements each with different data types.

Schemas in G-Portal are defined using XML Schema. Similar to the concept of a Class in the Object-Oriented model, schemas may inherit from other schemas. Having inheritance between schemas allows multiple schemas to share common attributes that can be used in classification. A Schema can also be part of another schema, a feature which allows us to compose new schemas from existing schemas. By allowing inheritance and composition of, we have a more flexible and expressive data model for defining complex geospatial and georeferenced resources. At the same time, we also enhance the reusability of schemas.

Figure 2 gives an example definition of a resource about flowers. Note that common attributes of a resource such as identifier, location, and access control information are defined by the base type *ResourceType* in the file Resource.xsd. *FlowerType* reuses these common attributes by including the file Resource.xsd and extending (subclassing) ResourceType. An example of an instance of the resource type Flower is shown in Figure 3. Location and access control information are not shown.

3.2 Resource Population

In G-Portal, resources are bundled into layers which then are organized into projects. Layers and projects are logical concepts that aim to help users better organize their geospatial and georeferenced resources and to perform their tasks more efficiently. This logical grouping also helps ease the visualization of resources which is important when the volume of resources becomes large. Visualization will be elaborated in Section 4.

To begin populating G-Portal with resources, a user who has administrator access rights has to first create a project using the Project Management Module. As discussed, a project is meant for a specific task that is related to geospatial or georeferenced data. A project has a unique identifier, a name, a general description and information for classification purposes. The creator of the project can determine the shareability of the project and this is also stored as part of the project information in the Project Database.

Before populating each project, the administrator must define the *target area* of the project which is by default the global atlas. The administrator then defines one or more *core layers* which will exist throughout the whole project as the spatial base for reference by other resources. For target areas other than the global atlas, the project administrator can create new core layers or replicate core layers from the project and layer database. For example, in the case of a small target area such as a project about Singapore, the effect of zooming in the global atlas would result in a monocolor and coarse Singapore map. The administrator might thus want to use a map of Singapore with a higher resolution as the core layer.

The separation of core layers from other layers is essential in G-Portal. The core layers are required in order to make G-Portal's concept of visualizing geospatial and georeferenced resources meaningful. The core layers typically consist of basic elements of a map such as country boundaries, lakes, rivers and their relevant information. For example, a project involving the study of various lakes in South-East Asia may combine the map of South-East Asian countries (one type of resource) and information about lakes (another type of resource) into one core layer. Alternatively, the project administrator may want to put these two types of resources into two core layers, since he/she may have created another project about South-East Asia and may want to replicate the core layer of the countries from that project. In this way, the administrator only needs to create the core layer for the lakes.

Layers that are not defined as core layers are called *non-core layers* or normal layers. These layers may or may not be shown on the map, depending on the preferences and interests of the individual end users. Non-core layers are defined by the project administrator and can contain any information under the topic of interest of the project. For example, in the South-East Asian lakes project, related resources might include rivers and mountains in the same area and these may form the non-core layers.

To define a layer, be it core or non-core, the administrator determines the types of resources that are to be included in the layer using the Layer Management Module. A unique identifier will also be assigned to the layer. In fact, what needs to be specified are the schema pointers of the resource, since schemas denotes the type or class of resources. The actual layer will then contain instances of the resources. Similar to projects, the access control information will be decided by the creator. Typically, core layers are configured as read-only while some of the non-core layers may be set to be modifiable by the users of the project or even the public.

To create a layer by combining different types of resources, the administrator needs to have the resource schemas at hand to select from. G-Portal's schema database maintains a large set of schemas that can be readily used. This collection is either provided by G-Portal as default resource schemas or accumulated from previous projects. If no suitable schema can be found, the administrator may create new schemas for the types of resources to be included in the project. Here, creating a new schema does not mean developing a schema from scratch. As mentioned in Section 3.1, the schema definition in G-Portal allows inheritance and composition. New schemas can be defined easily by "subclassing" existing schemas if a parent-children relationship is identified, or by including other schemas if a whole-part relationship is identified. The creation of new schemas or modification of existing schemas is facilitated by the Schema Management Module. The module also provides an interface for browsing the schema database to help users to locate appropriate schemas.

Once the required resource schemas and layers are defined, the project is ready for resource population by users. As mentioned earlier, authorized users will be able to contribute new resources into the project or update existing resources, and these tasks are performed through the Resource Management Module. The module is capable of understanding the syntactic structure of a resource by reading the resource's schema. The module will also help to enforce the correctness and integrity of the populated resources by validating the resource instances using the schema.

One of G-Portal's main objectives is to provide an environment for **user contribution** of geospatial and georeferenced resources – that is, the collection is not solely developed from G-Portal's administrators but from users of the digital library as well. In fact, we expect that most of the resources will be contributed through digital library users. G-Portal facilitates the process by providing modules that allow users to easily contribute resources and to help them to efficiently browse and retrieve the collection. In addition to supporting the manipulation of resources, G-Portal's management modules also allow users to create projects and layers and to define the schemas of the resources that they contribute. In this way, users contribute not only the collection but also the schemas of the contents. This is one of the features that distinguish G-Portal from other geospatial and georeferenced digital libraries (GDLs) such as DLESE [13] and ADL Gazetteer [6] which restrict resources to certain predefined schemas. Note however that to maintain the integrity of the collection, users' contributions are subject to review to ensure that certain standards are met. These will of course depend on the policies put in place by the administrators of G-Portal.

In the current implementation of G-Portal, each instance of a resource is stored as an individual XML document in an XML database. Schemas are stored in the XML database. Information about projects and layers are, on the other hand, stored in a relational database since they do not have complex tree structures as in schemas. User and access control information are also stored in the relational database.

```
<?xml version="1.0" encoding="UTF-8"?> <xsd:schema
    xmlns:xsd="http://www.w3.org/2000/10/XMLSchema"
    elementFormDefault="qualified">
  <xsd:include schemaLocation="Annotation.xsd"/>
  <xsd:complexType name="FlowerAnnotationType">
    <xsd:complexContent>
      <xsd:extension base="AnnotationType">
        <xsd:sequence>
          <xsd:element name="Name" type="xsd:string"/>
          <xsd:element name="Description" type="xsd:string"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>
  <xsd:element name="Resource">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="FlowerAnnotation"
                     type="FlowerAnnotationType"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Figure 4: A Resource Schema of Flower Annotation

```
<?xml version="1.0" encoding="UTF-8"?> <Resource
    xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="FlowerAnnotation.xsd">
  <FlowerAnnotation ID="3AX10M">
    <Location Type="Association">
      <AssociatedResource>
        <ID>12BCk_2</ID>
      </AssociatedResource>
    </Location>
    <Name>Rose annotation</Name>
    <Description>
      This is an annotation about this particular rose
    </Description>
  </FlowerAnnotation>
</Resource>
```

Figure 5: An Instance of Flower Annotation

3.3 Resource Annotation

In G-Portal, we deal not only with geospatial information, but also with georeferenced data. We introduce the concept of an **Annotation** to allow authorized users to attach information about existing resources.

Annotating existing resources is an important way of contributing to G-Portal and it is even possible to have annotations for existing annotations if required. Similar to other types of resources, annotations have their own schemas. Users may either select from our repository of schemas or define their own using the Schema Management Module. They can further contribute to G-Portal by submitting their schemas into the schema database.

Figure 4 shows the annotation schema definition. This schema extends the base schema AnnotationType which is derived from ResourceType. This relationship is possible because annotations are considered as a type of resource. An example of an annotation of this type is shown in Figure 5.

3.4 Access Control

Every resource, layer and project in G-Portal has access control information associated with it, and this information is defined by the resource creator.

In the current implementation of G-Portal, we employ a Unix-like security scheme. Each item (resource, layer, or project) in G-Portal has read and write flags for the three different categories of users: Owner, Group and Others. Each item will also have a creator and a group. The read and write flags are simple boolean values that indicate whether the operation is allowed. These flags can only be modified by the creator. Users of G-Portal may belong to one or more groups, and users in the same group as the creator are constrained by the read/write flags of the Group category. Users not in the same group as the creator will be considered to be in the Others group, and are restricted by the read/write flags of that group. Anonymous users do not belong to any group and are always considered as belonging to the Others group.

Finally, layers in one project can only be accessed within the project. If users of other projects want to reuse the layers in another project, they may replicate the layers into their own projects. This will likely to happen to layers that serve as core layers.

4. VISUALIZATION OF RESOURCES

In this section, we describe how G-Portal presents its resources to users. There are two basic interfaces for visualizing resources in G-Portal: a **Map-based Interface** and a **Classification Interface**.

4.1 Map-based Interface

In G-Portal, the predominant mode for accessing resources is via a map-based interface. This includes geospatial resources such as countries, rivers and mountains. As mentioned in Section 3, each project will have one or more core layers that contain only geospatial resources. Since each geospatial object has a *location* on the map and a *geographical representation* which is either a shape or an image, it is always possible to visualize the core layers on the map. The core layers will always be visible on the map and will appear at the lowest levels to serve as the geographical context of the whole project.

Non-core layers that contain geospatial objects will also be visualized in the same way. However, because these layers may not be essential to the project at any one time, they may be turned off by the users, i.e. made invisible. When a layer is turned off, all the instances of the resources grouped in the layer will no longer be visible on the map. Layers are ordered so that layers with higher z-indexes will overlay lower z-index layers. This ordering of layers can be predefined by project administrators and altered by users at the time of map access. The visibility of non-core layers can also be configured by the user.

When a resource does not have a geospatial representation so as to make it visualizable, symbolic icons may be used to represent it visually. Resources that fall into this category include georeferenced resources that are associated with certain instances of geospatial resources. In most cases, these resources are annotations and they automatically share the same geographical location as the geospatial resources that they are associated with. To display them, we will only need to give them some kind of visual representation which we achieve by assigning icons. Alternatively, users may choose not to display these resources on the map interface. Users can define in the schema how unvisualizable resources are displayed.

To allow users to navigate the map-based interface, G-Portal provides a set of *Navigation Tools*. For example, users can zoom in and out, and pan within any region of the project. G-Portal also provides a navigation aid called a "Bird Eye's View" that shows the entire region of the project in a small dialog and highlights the area that the user is currently viewing. All layers are listed in the *Layer Dialog* and users can turn the layers on and off using checkboxes. Layer colors may be changed here as well. The map-based interface together with these tools are shown in Figure 6.

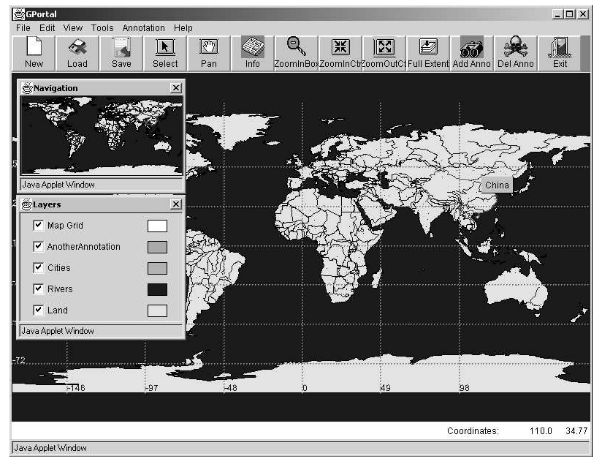

Figure 6: Screen Shot of G-Portal Client

4.2 Classification Interface

There exists another category of resources that are not associated with any specific instances of geospatial objects nor have they any geographical attributes. These include general geographical information such as how mountains are formed. Resources like climate and demographic information may also fall into this category.

G-Portal provides another type of visualization called the Classification Interface to present this category of resources. The Classification Interface categorizes and displays resources according to certain classification criteria and exists in parallel with the map-based interface.

The criteria for classification are usually based on some common attributes of the resources. The interface utilizes the classification service provided by G-Portal and displays the results in the form of hierarchical categories to users.

While non-geospatial resources can only be visualized in the Classification Interface, this interface is not restricted only to these resources. In general, any resource that has attributes that can be classified may be visualized using this interface. The classification interface is thus able to visualize all resources in G-Portal including those resources that are visualizable in the map-based interface.

4.3 Synchronization Between Interfaces

Besides providing support for visualizing resources in both interfaces, G-Portal also supports synchronization between the two interfaces. To illustrate this, we refer to the use case described in Section 1.1.

In the flora and fauna project, a user may create a schema for flowers found in Singapore. This schema may contain attributes such as species, color and locations found. While the flower resources may be classified by species or color, they can also be shown in the map-based interface according to their locations. By clicking on a flower resource in the classification interface, the user will see that the resource is also highlighted in the map interface.

In the classification interface, flower resources will also be grouped by different species and colors. For the color red, there may be several flower resources classified. If the user clicks on the category of red flowers, all the resources in that category will be shown on the map as well. On the other hand, if a user selects an area on the map that includes several flower resources, the corresponding resources will be highlighted in the classification interface. This synchronization will allow users to immediately see the distribution of resources according to different categories.

4.4 Viewing Resources

Viewing the actual content of a resource involves displaying both its geospatial and non-geospatial attributes. These attributes may not be easily visualized as the non-geospatial attributes can be of any data type, depending on the schema used. G-Portal offers the flexibility of visualizing resource contents by allowing users to specify the viewing mechanism within the schema itself. For example, external tools might be used if the attributes are more complex than G-Portal's built-in tools can handle.

4.5 Implementation

G-Portal's client-sides modules are currently implemented as a Java applet as this allows users to access the portal through web browsers. In addition, the Java programming language also provides expressive and powerful programming functionality which enables us to implement the features discussed in this paper. G-Portal is developed using Java 1.1 which is supported by the majority of the web browsers currently in use without the need for additional software. Figure 6 shows the main user interface of G-Portal.

The implementation of the map-based interface employs an open source software package called GeoTools [4], which provides a set of Application Programming Interfaces that encapsulate the common functionalities required by geographical-based systems. The management modules for manipulating projects, layers, resources and schemas are implemented in the Java applet as well.

5. RELATED WORK

G-Portal shares similar goals with existing digital libraries providing access to geospatial and georeferenced content. These include early systems such as Georep [10] and the Spatial Document Locator System (SDLS) [9] both of which provide basic search and retrieval services of geospatial data over the World Wide Web. More recent projects include the Alexandria Digital Library ADL [11], its successor the Alexandria Digital Earth Prototype System (ADEPT) [12], the Digital Library for Earth System Education (DLESE) [13] and Earthscape [3].

ADL's goal was to build a distributed digital library accessible over the Internet for geographically referenced materials including maps, satellite images, etc., and their associated metadata. ADEPT builds upon ADL and seeks to support the creation of personalized digital libraries of geospatial information ("learning spaces"), and investigating their utility in post-secondary science education [2]. Like ADEPT, DLESE focuses on education and contains online education resources for various educational levels in earth system science education. A notable difference is that the contents of the digital library relies on users' contribution of resources which may include maps, simulations, lesson plans, data sets, etc. Similarly, Earthscape provides a collection of online resources on the earth sciences and these are classified into four categories: teaching (e.g lesson plans), learning (e.g. readings and links), policy and research. Earthscape however differs from DLESE and G-Portal in that it is subscription-based with its resources obtained after review by an editorial board consisting of scholars in the discipline.

In addition, G-Portal shares the view that digital libraries should be environments where patrons not only retrieve information but also contribute resources to enhance the libraries' holdings as well. As such, G-Portal may be compared to digital libraries such as DLESE, Synchrony [5], the Global Digital Museum (GDM) [14] albeit with different approaches to contribution and in different domains. For example, DLESE and GDM support contributions via Web-based interfaces while Synchrony employs a spatial hypertext interface.

The G-Portal effort is also related to systems that attempt to provide a single interface to access multiple heterogenous GDLs. For example, System for the Optimized Selection of Spatial Data (SOS-SD) [8] aims to investigate the possibility of building a data warehouse application that retrieves preliminary data from different GDLs and integrate them into a single database for user inquiry. MEADOW (Middleware for Efficient Web-based Access to Databases through OpenGIS Wrappers) [1] adopts the CORBA object model and OpenGIS standard and proposes an architecture for accessing multiple geographic databases.

Finally, the approach of using map-based interfaces to access geospatial and georeferenced content may also be found in several systems. Apart from ADL and ADEPT, GlobalAtlas [7] is a geographical search service that allows users to search for distributed Web resources by drawing a query on a map. The system then responds by retrieving documents found within the bounded region. Similarly, the work by Zhao et al [16] proposes another map-based web search interface that integrates with the WWW and visual spatial query interface. SDLS also supports map-based access to documents although navigation and not querying is the main access mechanism. Users begin with a top-level map and traverse various regions ("information zones") until the desired documents are located. Note however that in these systems, search and retrieval are the main features whereas G-Portal focuses on resource contribution by users as well.

Despite some similarities with existing GDL projects, G-Portal has several important differences. For example, both ADL and ADEPT own the resources in the collection while G-Portal adopts a model closer to DLESE in that the development of the collection depends mainly on users' contributions as well as on the discovery and acquisition of external resources (such as geography-related Web sites). On the other hand, G-Portal's resource access and contribution approach differs from DLESE and Earthscape through the provision of both an interactive map-based interface as well as a Web-based interface. Either interface may be used depending on the type of resource. For instance, spatial information may be accessed through the map-based interface while non-spatial resources may be access through the Web-based interface. This flexibility allows a greater range of resources and resource types to be made available to users of the digital library.

6. CONCLUSION

In the G-Portal project, a digital library of geospatial and georeferenced resources is constructed and organized as a unique web portal capable of presenting information both spatially in a map-based interface and non-spatially in a hierarchical classification interface. With G-Portal, users can not only view these resources but also contribute geospatial and georeferenced web resources under different projects using pre-defined XML schemas. By allowing resources to be shared across projects, G-Portal facilitates sharing of web resources while ensuring that they are compliant with predefined schema structures useful in visualization and query formulation. G-Portal also adopts generic functional modules around its resource repositories making it easier to support different use cases for learning and research. In this paper, we described the system architecture of G-Portal and its modules. Its resource management and visualization capabilities are also discussed in detail.

Work on G-Portal is ongoing with the project focusing on the following tasks:

- *Implementation of G-Portal:* G-Portal is a 6-month old project and while most of the design work has been completed, the implementation of a working prototype was started only recently. The preliminary versions of the visualization and resource management modules have been completed but there are several features yet to be implemented, e.g. the classification interface and the synchronization between the map and classification interfaces. The other modules, e.g. query engine and web resource maintenance are currently being developed.

- *Evaluation of G-Portal:* The success of G-Portal depends on how well it can be used to support the different use cases for learning and research involving geospatial and georeferenced web content. Plans are underway to conduct experiments in Singapore schools and other interested organizations to evaluate the use of G-Portal to support such use cases.

- *Remote resource query processing:* Much of the geospatial and georeferenced web information are currently stored in databases and can only be accessed via simple query interfaces. The heterogeneities among these interfaces pose challenges to G-Portal and work is being done to provide an unified query interface to these systems while keeping the query evaluation process efficient.

7. ACKNOWLEDGMENTS

This work is funded by the SingAREN Project M48020004.

We would also like to thank Terrence Smith, Linda Hill, and Christine Borgman for their advice during the initial stage of this research.

8. REFERENCES

[1] S. K. Cha, K. Kim, C. Song, Y. Kwon, and S. Hwang. Efficient web-based access to multiple geographic databases through automatically generated wrappers. In *Proceedings of the First International Conference on Web Information Systems Engineering (WISE 2000)*, pages 34–41, Hong Kong, China, June 2000. IEEE Computer Society Press.

[2] A. Coleman, T. Smith, O. Buchel, and R. Mayer. Learning spaces in digital libraries. In *Proceedings of Fifth European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, pages 251–262, Darmstadt, Germany, September 2001. Berlin: Springerf.

[3] Columbia Earthscape. http://www.earthscape.org.

[4] GeoTools Homepage. http://geotools.sourceforge.net.

[5] D. Goh and J. Leggett. Patron augmented digital libraries. In *Proceedings of the Fifth ACM Conference on Digital Libraries (DL'00)*, pages 153–163, San Antonio, Texas, USA, June 2000. ACM Press.

[6] L. L. Hill. Core elements of digital gazetteers: placenames, categories, and footprints. In J. Borbinha and T. Bakerf, editors, *Proceedings of Fourth European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2000)*, pages 280–290, Lisbon, Portugal, September 2000. Berlin: Springerf.

[7] F. Lee, S. Bressan, and B. Ooi. Hybrid transformation for indexing and searching Web documents in the cartographic paradigm. *Information Systems*, 26(2):75–92, 2001.

[8] F. Létourneau, Y. Bédard, and M.-J. Proulx. SOS-SD a data warehouse-based system for the optimized selectinof spatial data. D-Lib Magazine, March 1997.

[9] J. Orendorf and C. Kacmar. A spatial approach to organizing and locating digital libraries and their content. In *Proceedings of the First ACM Conference on Digital Libraries (DL'96)*, pages 83–89, Bethesda, MD, USA, March 1996. ACM Press.

[10] M.-J. Proulx, Y. Bédard, F. Létourneau, and C. Martel. GEOREP: A WWW customizable georeferenced digital library for spatial data. D-Lib Magazine, December 1996.

[11] T. Smith. A digital library for geographically referenced materials. *IEEE Computer*, 29(5):54–60, 1996.

[12] T. Smith, G. Janee, J. Frew, and A. Coleman. The Alexandria Digital Earth ProtoType system. In *Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001)*, pages 118–119, Roanoke, VA, USA, June 2001.

[13] T. Sumner and M. Dawe. Looking at digital library usability from a reuse perspective. In *Proceedings of the First ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001)*, pages 416–425, Roanoke, VA, USA, June 2001.

[14] J. Takahashi, T. Kushida, J. Hong, S. Sugita, Y. Kurita, R. Rieger, W. Martin, G. Gay, J. Reeve, and R. Loverance. Global digital museum: Multimedia information access and creation on the internet. In *Proceedings of the Third ACM Conference on Digital Libraries (DL'98)*, pages 244–253, Pittsburgh, PA, USA, June 1998. ACM Press.

[15] Yahoo! http://www.yahoo.com.

[16] X. Zhou, J. D. Yates, and G. Chen. Using visual spatial search interface for WWW applications. *Information Systems*, 26(2):61–74, 2001.