Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

9-2005

On organizing and accessing geospatial and georeferenced web resources using the g-portal system

Zehua LIU

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

Yin-Leng THENG

Dion Hoe-Lian GOH
Nanyang Technological University

Wee-Keong NG

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, and the Numerical Analysis and Scientific Computing Commons

Citation

On organizing and accessing geospatial and georeferenced Web resources using the G-Portal system

Ee-Peng Lim [a,*], Zehua Liu [a], Ming Yin [a], Dion Hoe-Lian Goh [b], Yin-Leng Theng [b], Wee Keong Ng [a]

[a] *Centre for Advanced Information Systems, School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore*
[b] *School of Communications and Information, Nanyang Technological University, Singapore 639798, Singapore*

Abstract

In order to organise and manage geospatial and georeferenced information on the Web making them convenient for searching and browsing, a digital portal known as G-Portal has been designed and implemented. Compared to other digital libraries, G-Portal is unique for several of its features. It maintains metadata resources in XML with flexible resource schemas. Logical groupings of metadata resources as projects and layers are possible to allow the entire metadata collection to be partitioned differently for users with different information needs. These metadata resources can be displayed in both the classification-based and map-based interfaces provided by G-Portal. G-Portal further incorporates both a query module and an annotation module for users to search metadata and to create additional knowledge for sharing respectively. G-Portal also includes a resource classification module that categorizes resources into one or more hierarchical category trees based on user-defined classification schemas. This paper gives an overview of the G-Portal design and implementation. The portal features will be illustrated using a collection of high school geography examination-related resources.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Digital library portal; Geospatial and georeferenced resources; Metadata management

---

* Corresponding author.
  E-mail addresses: aseplim@ntu.edu.sg (E.-P. Lim), liuzh@pmail.ntu.edu.sg (Z. Liu), asmyin@ntu.edu.sg (M. Yin), ashlgoh@ntu.edu.sg (D.H.-L. Goh), tyltheng@ntu.edu.sg (Y.-L. Theng), awkng@ntu.edu.sg (W.K. Ng).

1. Introduction

In this paper, we describe a open digital library system known as G-Portal that organizes geospatial and georeferenced Web resources, and supports both navigational and query access on its resources. The design of G-Portal is very similar to Web portals such as Yahoo! [1] where selected Web resources are indexed into a large hierarchy of categories suitable for searching and browsing. The Web portal concept offers several known advantages to their users. Firstly, 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 search engines to query their indexed resources.

To manage geospatial and georeferenced Web resources, G-Portal extends the traditional Web portal concept with services that cater for resources with spatial features including location and geometry information. As Web users are likely to locate geospatial resources based on their spatial features, G-Portal provides a map-based interface to display information, and integrates spatial features into its query interface. To meet the information needs of different user communities, G-Portal organizes resources into different projects and layers. Based on the resource schemas and user needs, different classification schemas can be defined to support alternative ways to classify and visualize the resources. Furthermore, to effectively use G-Portal in geography education, an annotation subsystem is included to allow users to annotate resources as they explore the collection. Annotations are treated as a special kind of resources, and hence they could also be accessed using the same visualization and query facilities designed for other resources. By making annotations available among users within a group, knowledge sharing becomes possible over G-Portal.

The design of G-Portal covers a wide range of research issues. In this paper, we will confine our discussion to those issues related to metadata resource representation and organization including the classification structures and user interface presentation. Other issues such as query and user access control issues will only be covered briefly as they can be treated quite independently. As G-Portal is designed to support learning, we have also conducted some study on its use in some learning/pedagogical tasks. Interested readers can refer to Chua, Goh, Lim, Liu, and Ang (2002) and Theng et al. (2002) for more information.

1.1. A use case example

In our design process, we conducted a use-case analysis (Fowler & Scott, 1999) on how G-Portal should be used to meet the learning goals of its users. Each use case describes the steps required to accomplish a task using geospatial and georeferenced resource collection(s). The development of such use cases for earth system education has been carried out in the DLESE (Digital Library for Earth System Education) research (Sumner & Dawe, 2001). Instead of using some of the use cases from DLESE, we have decided to examine a use case most pertinent to the learning needs of high school students in Singapore, that is, preparing for the nation-wide Geography examination. This use case can guide us in determining specific G-Portal design requirements, and at the same time, be general enough to be used for other learning applications.

High school equivalent of students in Singapore in their fourth year are required to take a nation-wide examination covering 6–8 different subjects, including Geography. Geography is not an easy nor interesting subject for the students as they usually perceive the subject to be very uninteresting and boring. Moreover, as the students' advancement to the tertiary program heavily depends on the outcome of the examinations, the teaching of Geography usually follows the examination syllabus very closely, and the preparation for the Geography examination heavily relies on past year examination questions and answers offered by local publishers. To correct this narrow approach to study geography, we intend to develop a digital library

---

[1] http://www.yahoo.com

portal system to help students expand their knowledge beyond the "model" answers of examination questions by exploring knowledge from other sources, particularly on the Web.

A Geography teacher (or a group of teachers) can first compile a collection of past year examination questions and other related questions, known as the examination question resource collection. Using G-Portal, this collection can be browsed and searched by teachers and students. There are two kinds of information that can be gathered for each question, namely suggested answers and knowledge about concepts referenced by the question. Students and teachers can work together to gather such information, and also provide comments. The examination questions, answers and other resources can be arranged in different ways for users to locate them easily based on their types and attributes. Many of them also involve spatial information that can be best visualized using a map. When the resources are existing Web sites or Web pages, it is necessary to support links to these resources from G-Portal.

The above use case is a fairly general and open-ended one. It involves some general tasks to be performed on geospatial and georeferenced Web resources but does not dictate a strict ordering on the tasks. Instead of trying to address the above requirements by a single workflow, we have decided to have G-Portal provide a set of services that can be combined in a flexible way to achieve the use case requirements.

From the use case example, we also conclude that G-Portal should support a flexible repository system accommodating different kinds of resources. The knowledge managed by G-Portal can be classified broadly under general resources and annotations. The former refers to the core knowledge for users to explore and learn, while the latter refers to additional contributed knowledge about the former. The format of the general resource and annotation description should be made common in order for G-Portal's services to be developed in a consistent way. Nevertheless, the format should also allow users to distinguish between the types of information they deal with. One should also be able to group different resources and annotations together under a common project. Furthermore, to facilitate searching and browsing, alternative category structures should be made available to users to suit different information needs and contexts.

The above suggests the need for a set of modules to deal with the different services. It turns out that modules can be designed to be generic enough to handle other use cases, such as collecting and organizing information for a Geography project, and establishing a special interest group for learning specific Geography topics.

1.2. Key concepts

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

Resources are the basic elements of G-Portal. They are metadata describing raw resources, for example, physical books, journal articles, papers, Web pages, and Web sites. They are generic containers that store data. Every resource has an *identifier*, a *resource name*, a *location*, a *creator*, *source,* and *content* as defined in its schema. *Annotations* are a special type of resources that are associated with some other resources. Annotations are typically used by users to attach their knowledge about resources and share this knowledge with other users. In G-Portal, all resources, including annotations, 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 paradigm.

In addition, resources typically are not used in isolation but as part of a larger task. The concept of a *Project* is introduced to define the collection of resources (and annotations) that are relevant to a use case, for example, geography examination related resources. Under a project, resources are further grouped into *layers* for finer grained organization. The display of all resources within a layer can be turned on and off within the G-Portal user interface to reduce information overload. Each project will have one or more *core layers*, which serve as the spatial context for the entire project.

The separation of core layers from other layers is essential in G-Portal. The core layers typically consist of resources which are spatial objects 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. Layers that are not defined as core layers are called *non-core layers*. Core layers will always be shown on the map-based interface. The display of non-core layers on the map can be toggled on or off, depending on the preferences and interests of the individual end users.

1.3. Paper outline

The remaining sections of this paper present G-Portal in detail. We first review the existing works related to G-Portal in Section 2. The management of metadata resources and annotations is described in Section 3. Section 4 discusses how the system presents and visualizes the resources. Section 5 covers the mechanism for creating and editing annotation and metadata resources. The classification scheme is then presented in Section 6. Finally, we give some concluding remarks and directions of our future work in Section 7.

2. Related work

Compared to the previous digital library systems, G-Portal is unique in the way resources are represented and organized. The flexible use of resource schemas allow G-Portal to operate on a collection of resources of different types. G-Portal further supports configurable classification schemes for these resources making the user interface more accustomed to the learning task at hand.

G-Portal shares similar goals with existing digital libraries providing access to geospatial and georeferenced content. These include early systems such as Georep (Proulx, Bédard, Létourneau, & Martel, 1996) and the Spatial Document Locator System (SDLS) (Orendorf & Kacmar, 1996) 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) (Smith, 1996), its successor the Alexandria Digital Earth Prototype System (ADEPT) (Smith, Janee, Frew, & Coleman, 2001), the Digital Library for Earth System Education (DLESE) (Sumner & Dawe, 2001), Earthscape (Columbia Earthscape, 2002), and the Building the Digital Earth Project (BDE) (Seber, Sandvol, Brindisi, & Barazangi, 2001).

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 (Coleman, Smith, Buchel, & Mayer, 2001). 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 content of DLESE 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. The Building the Digital Earth Project, unlike G-Portal, focuses more on geospatial content creation and digital earth visualization (Seber et al., 2001).

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 (Goh & Leggett, 2000), the Global Digital Museum (GDM) (Takahashi et al., 1998) 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.

Despite some similarities with existing georeferenced and/or geospatial digital library (GDL) projects, G-Portal has several important differences. Its design allows flexibility in metadata resource representation and provides a generic approach to organise and manipulate these resources. 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). 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 classification-based interface. Either interface may be used depending on the type of resources. For instance, spatial information may be accessed through the map-based interface while non-spatial resources may be accessed through the classification-based interface. This flexibility allows a greater range of resources and resource types to be made available to the Digital Library (DL) users.

3. Modeling resources and annotations

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

3.1. Resource data model

In G-Portal, we distinguish the resources maintained by the DL system from those pre-existing Web resources located on public domain Web sites. The former are referred to as *metadata resources*, while the latter are known as *raw resources*. Raw resources could be webpages, maps, gazetteer (a dictionary of entities), etc., so long as they are related to geography and are accessible on the Web. Metadata resources are the more structured versions of their raw counterparts and are contributed by users knowledgeable about the associated Web sites or pages. G-Portal is designed to manage the metadata resources only. To allow for a more flexible use of these metadata resources, we also allow metadata resources to be created without associating them with Web resources. These are usually metadata resources that do not have raw resources readily available on the Web, for example, geography book metadata. The metadata resources in G-Portal are represented in XML format while the raw resources can be of any format. To simplify the rest of our discussion, the term "resource" refers to "metadata resource" unless in situations where explicit terms are required.

Each resource must be created using some *resource schema* denned using XML Schema. [2] All resource schemas are derived from a *base resource schema*, which is partially shown in Fig. 1. This base resource schema includes the common attributes of a resource such as identifier, name, location, creator, and source. Each resource is assigned a unique identifier within the G-Portal system. The location attribute registers the spatial properties of the resource that can be a point, line, polygon, or a set of polygons. The source attribute captures the link to the raw resource referenced by the metadata resource. Unlike all the other elements, the content element in the base resource schema does not have a fixed type. The content element structure can be customized in different derived resource schema to represent different types of resource content. [3] Since its content element type is not specified, the base resource schema serves as a virtual class for other *derived resource schemas* to be defined.

Each derived resource schema may include other attributes relevant to the type of resources described by the schema. These additional attributes are defined by overriding the ContentType element. For example,

---

[2] http://www.w3.org/xml/schema
[3] An alternative approach is to use namespace to distinguish metadata and content elements in a resource. This unfortunately is not supported in the current version of G-Portal.

```
<!-- Resource.xsd -->

<?xml version="1.0" encoding="UTF-8"?>

<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">

    <xsd:element name="Resource" type="ResourceType">

    <xsd:complexType name="ResourceType">

        <xsd:sequence>

            <xsd:element name="ID" type="xsd:string"/>

            <xsd:element name="ResourceName" type="ResourceNameType"/>

            <xsd:element name="Location" type="LocationType"/>

            <xsd:element name="Creator" type="CreatorType"/>

            <xsd:element name="Source" type="SourceType"/>

            <xsd:element name="Content" type="ContentType"/>

        </xsd:sequence>

    </xsd:complexType>

    ... ...

    <xsd:complexType name="ContentType">

    </xsd:complexType>

</xsd:schema>
```

Fig. 1. The base schema for all resource schemas.

one can define a schema for resources that describe information about high school geography exam questions. In this case, each question resource consists of the year, theme and topic defined in the syllabus, question number, the content of the question and optionally the choices if it is a multiple choice question. Since XML Schema is itself an XML document, resource schemas share advantages of all XML documents, including openness and machine-readability.

Resource instances can be defined based on actual values of resource attributes. Fig. 2 presents an example of a resource of the type ExamQuestion based on the ExamQuestion resource schema.

It should be noted that for this particular metadata resource, we have included some contents of the raw resource, in this case the content of the actual question, so that searching of questions can be more meaningful. In general, this needs not be the case. What attributes to be included in the metadata and whether the attributes come from the raw resource are completely determined by the metadata resource schema creators.

Another point to note is that the semantics of the elements, especially those under the Content element, may sometimes be ambiguous. This problem may be partially solved by introducing namespace. However, it might still be difficult to disambiguate the semantics unless there is an authority or some standard committee that defines the possible meaning of the elements in a registry. We also note that incorporating namespace into the system is rather straightforward. We only need to add the namespaces into the schema and the corresponding resources. There are no changes required in terms of implementation.

3.2. Annotations as metadata resources

To create an entire learning environment that allows the sharing of knowledge among users, it is often desirable to allow users to annotate resources with their knowledge, store these annotations in the DL and make them available to other users (Xerox, 2000).

```
<!-- ExamQuestionD87P1Q23.xml -->

<?xml version="1.0" encoding="UTF-8"?>

<Resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:noNamespaceSchemaLocation="ExamQuestion.xsd">

  <ID>ExamQuestion_1</ID>

  <ResourceName> <Name>D87P1Q23</Name> </ResourceName>

  <Location>... ...</Location>

  <Creator>... ...</Creator>

  <Source> <Link>
      http://www.cais.ntu.edu.sg/~gportal/GPortal/gce_exam/D87P1Q23.htm
  </Link> </Source>

  <Content>

    <Paper> <Year>1987</Year> <Month>12</Month> <Number>1</Number> </Paper>

    <Subject>

      <Title>Identify type of climate based on shaded regions on the map</Title>

      <Part>Physical Geography</Part>

      <Theme>Weather and Climate</Theme>

      <Topic>Climatic types: Nature, distribution and explanation</Topic>

    </Subject>

    <QuestionNumber>23</QuestionNumber>

    <Type>MCQ</Type>

    <Description>The world map shows the distribution of one type of climate.
             This type of climate is described as </Description>

    <MCQChoices>

      <Choice> <Label>A</Label> <Description>equatorial</Description> </Choice>

      <Choice> <Label>B</Label> <Description>hot desert</Description> </Choice>

      <Choice> <Label>C</Label> <Description>savanna</Description> </Choice>

      <Choice> <Label>D</Label> <Description>tropical monsoon</Description> </Choice>

    </MCQChoices>

  </Content>

</Resource>
```

Fig. 2. An example of an ExamQuestion resource.

In the area of Web annotation, the Annotea Web-based annotation system supports a RDF-based annotations of Web pages or other Web objects with URIs (Universal Resource Identifiers) (Kahan & Koivunen, 2001). Each annotation is represented as a set of elements (about the annotation creator, date of creation, etc.) and an annotation body. Multiple annotation classes (similar to our annotation schemas) can be created by Annotea to instantiate different annotations for different purposes (e.g. Advice,

Comment, Example, etc.). However, the content or body of annotations in Annotea is not structured (free text or HTML document). The approach of using unstructured annotation content is also adopted in the DLESE and ADEPT digital library projects. DLESE and ADEPT define a common metadata framework known as ADN Framework to standardize the representation of metadata resources including annotations (ADN, 2003; Sumner & Dawe, 2001; Smith et al., 2001).

In G-Portal, annotations are associated with metadata resources that are proxies of Web resources. Unlike the systems mentioned above, G-Portal has adopted a more flexible annotation schema structure which allows a basic annotation schema to be extended with different elements that capture a wide variety of annotation types. Annotations are treated as a special type of resource (Liu, Lim, & Goh, 2002). In addition to the elements of the basic resource schema, a new `AnnotatedResources` element is added to the base annotation resource schema as shown in Fig. 3. This element identifies one or more resources that have been annotated by keeping the *resource ids* of the list of annotated resources. Other important elements, such as id and location elements, are inherited from the basic resource schema.

As a subtype of resource, annotations can enjoy the same DL services as other ordinary resources (i.e., those resources that are not annotations). Furthermore, annotations can also be annotated, just like ordinary resources. Similarly, annotations within a project are grouped into different layers. This design option greatly simplifies the implementation of G-Portal.

Another side effect of an annotation being a resource is that every annotation will have a location element. This location element can be derived from the resources being annotated, if they have geospatial attributes. An annotation may also have its own location value directly assigned by the annotator.

```
<!-- BasicAnnotation.xsd -->
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
    <xsd:redefine schemaLocation="Resource.xsd"/>
    <xsd:element name="Resource" type="AnnotationType"/>
    <xsd:complexType name="AnnotationType">
        <xsd:complexContent>
            <xsd:extension base="ResourceType">
                <xsd:sequence>
                    <xsd:element name="AnnotatedResources" type="AnnotatedResourcesType"/>
                </xsd:sequence>
            </xsd:extension>
        </xsd:complexContent>
    </xsd:complexType>
    <xsd:complexType name="AnnotatedResourcesType">
        <xsd:sequence>
            <xsd:element name="Resource" type="xsd:string" maxOccurs="unbounded"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

Fig. 3. The base schema for annotations.

Unlike most existing annotation frameworks where annotations can only be created for individual resources, our framework allows multiple resources (possibly with different resources schemas) to be annotated using a single annotation. In the geography exam resource example, an annotation may be created

```
<!-- QuestionAnnotation.xsd -->
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
    <xsd:redefine schemaLocation="BasicAnnotation.xsd">
        <xsd:complexType name="ContentType">
            <xsd:sequence>
                <xsd:element name="Comment" type="xsd:string"/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:redefine>
</xsd:schema>
```

Fig. 4. A simple schema of annotations on exam questions.

```
<!-- TrendQuestionAnnotation1.xml -->
<?xml version="1.0" encoding="UTF-8"?>
<Resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:noNamespaceSchemaLocation="QuestionAnnotation.xsd">
    <ID>QuestionAnnotation_1</ID>
    <ResourceName>
        <Name>Trend in vegetation related questions</Name>
    </ResourceName>
    <Location Type="Geometry">  <Geometry> ... ... </Geometry>  </Location>
    <Creator> ... ... </Creator>
    <Source> ... ... </Source>
    <Content>
        <Comment>A question on vegetation seems to appear each year!</Comment>
    </Content>
    <AnnotatedResources>
        <Resource>ExamQuestion_35</Resource>
        <Resource>ExamQuestion_60</Resource>
        <Resource>ExamQuestion_61</Resource>
    </AnnotatedResources>
</Resource>
```

Fig. 5. A simple annotation on exam questions.

to describe a geography concept that appears in multiple examination question resources. In traditional annotation systems, this can only be achieved by annotating one question resource and adding the ids of the other question resources into the annotation content, which is a non-intuitive approach because an extra step is required to find all the annotated questions.

The definition of the content in the basic annotation schema is left to the users, just as in the definition of resources. New types of annotations can be created by altering the format of the content of the annotation. Fig. 4 shows an example annotation schema that has only one content element Comment. An instance of annotation of this type is shown in Fig. 5. This annotation captures a simple comment on the three instances of resources of the examination question type. The ids of the annotated resources (questions) are listed in the `AnnotatedResources` element. Here, only a simple comment is used to describe the trend spotted from these questions.

4. Map-based and classification-based user interfaces

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-based Interface*.

4.1. Map-based interface

The map-based interface is a window in which geospatial metadata resources, containing location and shape values in the location elements, can be displayed on a two-dimensional map.

As metadata resources are grouped under different projects for different application usages, the map-based interface (and the classification-based interface) is designed to display resources of one project at a time. Each project will have one or more core layers that contain geospatial resources to provide the map context for displaying resources in the non-core layers. For example, the core layers usually contain resources corresponding to country boundaries, rivers, mountains, etc., within a map. Hence, the resources in the core layers are always visible in the map interface while those in the non-core layers are optionally displayed as configured by the users.

Geospatial resources of the same layer are displayed using the same line and fill colors selected by the users. When it is necessary to have all resources of the same resource schema use the same color scheme, we can group the resources into a layer and define a color scheme for the layer.

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's Eye 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 Fig. 6.

4.2. Classification-based interface

G-Portal also manages non-geospatial resources that are relevant to a project. Examples of these resources include general geographical information such as how mountains are formed, climate and demographic information.

For these resources as well as the geospatial resources, G-Portal provides a Classification-based Interface that shows resources in categories defined by users, with categories organized into category trees. This classification-based interface exists side-by-side with the map-based interface to provide an alternative category-based approach to browse and search resources. The criteria for classification are usually based on
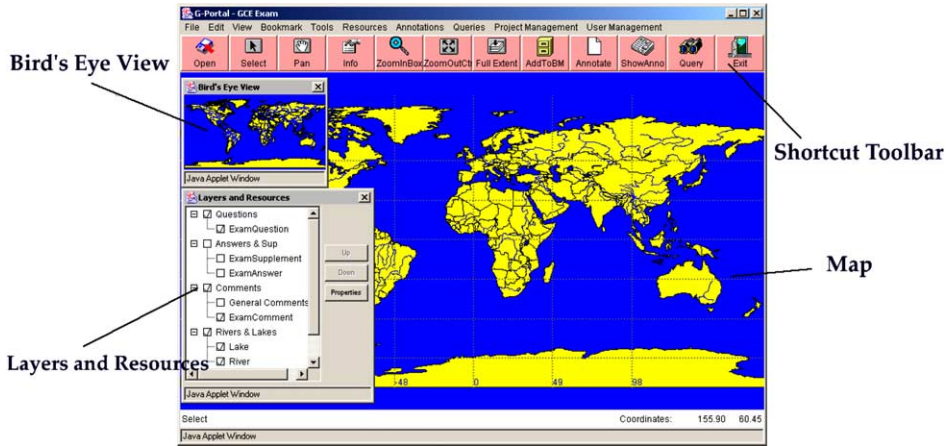
Fig. 6. Screen shot of G-Portal's map-based interface.

attributes of the resources. The interface utilizes the classification service provided by G-Portal and displays the results in the form of an hierarchy to users. In Section 6, we will describe the classification language used to define categories and the hierarchical category trees. Other than the mode of visualization, the map-based and classification-based interfaces are quite similar in terms of resource selection and display.

To the users, the same resources can be viewed under different category trees according to the grouping criteria most appropriate to the task at hand. For example, for students who want to analyse examination trends will prefer a category tree with question resources classified by year. For students who want to revise what they have learnt in class, question resources classified by topic may be more appropriate.

As shown in Fig. 7, the classification-based interface displays a single root that contains resource schemas as their child categories. For each resource schema, one can have one or more category tree structures each containing other internal categories and leaf categories. All resources appear at the leaf categories.
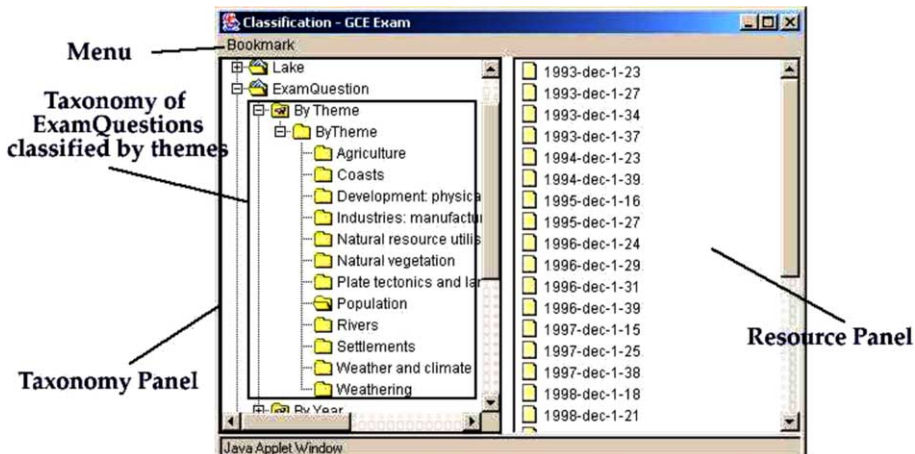

Fig. 7. Screen shot of classification interface.

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.

Assume that the geography examination question resources have been created for the geography examination use case. While the question resources may be classified by year or topic, they can also be shown in the map-based interface according to their locations. By clicking on a question resource in the classification interface, the user will see that the resource is also highlighted in the map interface if the question carries location information.

Suppose that the examination question resources are grouped by year. If the user clicks on the category of questions appearing in 2000, all the questions with spatial features 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 question resources, the corresponding resources will be highlighted in the classification interface. This synchronization will allow users to immediately see the distribution of resources in both interfaces, and to quickly retrieve relevant resources using alternative approaches.

4.4. Query interface

Besides the navigation-based approaches to locate resources using the two interfaces, G-Portal also offers a simple query interface for users to search resources within a given project. The resources to be searched and the search criteria are specified using the query user interface (see Fig. 8). To cater to resources with spatial features, the query user interface allows a query box to be drawn within the map-based interface and supports several basic spatial query predicates such as containment and overlap. Internally, the search criteria are converted into a XQuery-like query language known as $\mathcal{R2L}$ (Liu, Lim, Ng, & Goh, 2003) and submitted to the query engine at the G-Portal server for evaluation. The search results are shown in a result list window where resources can be selected for further viewing. A similar synchronization mechanism has been implemented
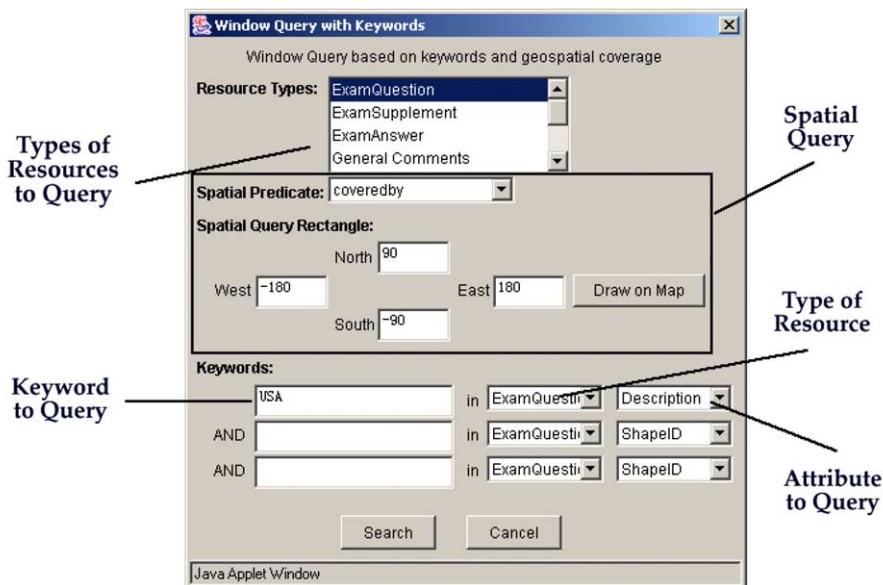

Fig. 8. The window query dialog.

between the result list window and both the map-based and classification-based interfaces. When the user selects a resource in the result list window, the location of the resource on the map and the categories that the resource is assigned to in the classification-based interface will also be shown.

5. Metadata resource creation

5.1. Resource/annotation creation interface

G-Portal provides a user interface to allow users to create resources and annotations. Creating a new resource instance consists of two main steps: specifying the core attributes and the customized attributes (i.e., those under the Content element). This subsection discusses the user interface for specifying core attributes. The interface for specifying customized attributes is covered in next subsection.

Fig. 9 shows the interface for creating a new annotation. [4] To start, the user chooses a schema for the new annotation from a list. The layer where the annotation belongs to is also specified. Once the schema and the layer are determined, the user can start entering the core attributes, such as name and source link. Note that users do not need to enter the resource id as it is automatically generated by the system. If the current user is using a public or shared account, he/she can change the author information to reflect his/her identity.

Another important core attribute is the location. For an annotation, there are several ways of assigning a location. It can be either derived from the location values of the annotated resources (using a minimum bounding box covering them, or simply using the union of their location values), or explicitly specified by the annotator using a drawing tool within the map interface. However, for a normal resource, the location can only be assigned using the drawing tool.

One important difference in the creation of resources and annotations is that creating an annotation requires the user to first specify the resources to be annotated. This can be performed easily by first bookmarking the resources to be annotated and specifying the annotation content. The bookmarking step essentially involves selecting resources from the map-based or classification-based interface and adding them to a bookmark one at a time. Once a set of resources are bookmarked, the user can start specifying the core attributes with the procedure outlined above.

5.2. Metadata resource editor

In the interface for creating annotation, after specifying all the core attributes, the user clicks on the button "Edit XML Content" to edit the customized attributes. Recall that all customized attributes must be defined under the Content element and may be arbitrarily nested. The editing of customized attributes, therefore, requires an editor capable of editing XML content. Since users may not be familiar with XML, it will be too difficult for them to use a simple text editor to enter XML content in plain text format. To make the metadata resource creation process much easier, we have implemented a user-friendly, schema-guided XML editor. This editor is used to edit both resource and annotation content.

Fig. 10 shows a screenshot of the XML editor for creating a question resource, with a schema described in Section 3.1. The appearance of the editor resembles that of Windows Explorer. The tree formed by the XML element structure is displayed in the left panel. Element names are shown as names of tree nodes.

---

[4] The interface for creating a new resource is very similar to that for annotation, except that the portion for specifying geospatial attribute is simpler.
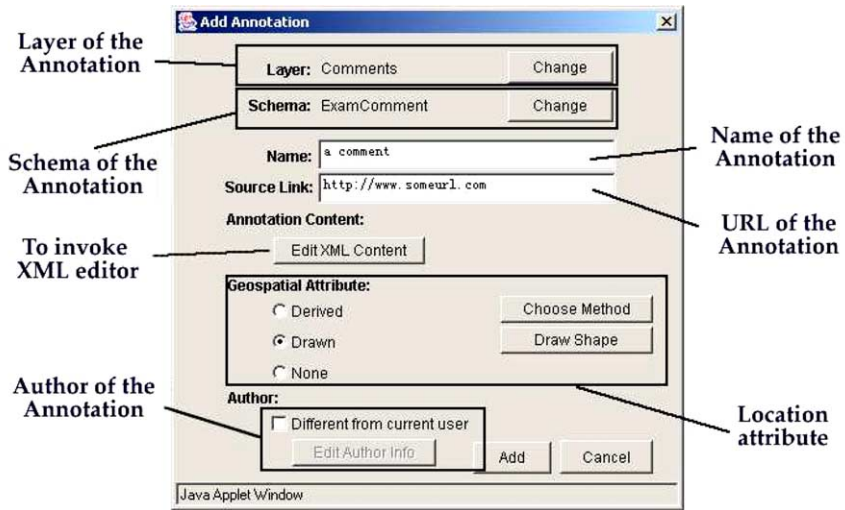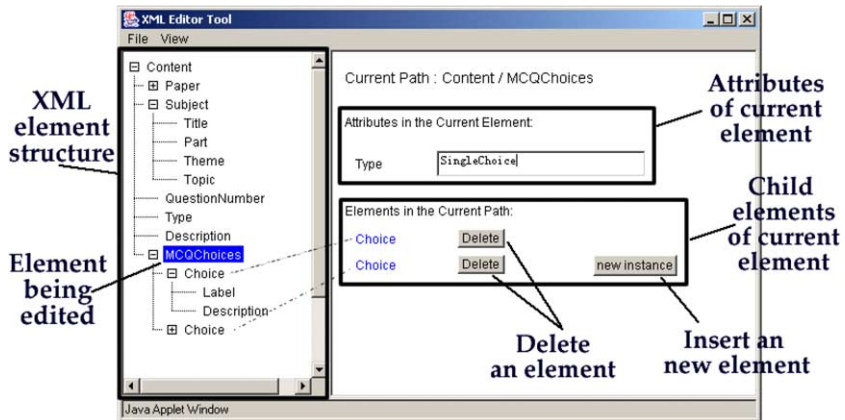
Fig. 9. The create annotation dialog.

Fig. 10. Screenshot of the XML editor.

When an element is selected, the right panel shows the attributes, text values and/or the child elements of the selected element.

One important feature of the XML editor is that it is schema guided. Before the editing starts, the resource schema (in the form of XML Schema) has to be supplied to the editor. The editor parses the schema and generates a base skeleton of the XML content with empty value for each element. The user then interactively uses the tree in the left panel to select the elements and fills in their text values and attribute values in the right panel.

In the screenshot in Fig. 10, the MCQChoices element (selected in the left panel) is being edited. As defined in the Question resource schema, a MCQ-Choices element, being of the complex type QuestionMCQChoice-Type, has one attribute named Type and may have one or more child elements named Choice. The possibility of one or more occurrences of the Choice element is indicated by the maxOccurs = "unbounded" attribute in the definition. The value of the attribute ("SingleChoice" in this case) can be filled in. For child elements

such as Choice where multiple occurrences are allowed, two operations can be performed: *insert* and *delete*. The insert operation is invoked by clicking on the "new instance" button, which results in one more instance of the element being created. The delete operation is performed by clicking on the "delete" button to the right of the element to be deleted.

Note that the availability of the insert and delete operations are strictly dependent on the given resource schema. For example, suppose that the question resource schema indicates the Choice element should have at least one occurrence. Thus, if in Fig. 10, the user deletes one of the two Choice elements, leaving only one Choice element left, the corresponding "delete" button of the Choice element left will not be shown so as to ensure that the user has no way to delete this last instance. Similarly, if the schema specifies that the element has exactly one occurrence, then neither the "delete" nor the "new instance" buttons will appear. Therefore, the insertion and deletion of elements is closely guided by the schema and the editor is able to ensure structural conformance at run-time. Such tight control over the allowed operations greatly reduces the chances of making errors; thus making the XML content creation process much easier for ordinary users.

Once the customized attributes have been specified, the XML editor verifies that the entered values satisfy all the constraints in the schema before saving the edited XML content and returning to the create resource/annotation interface. The complete resource/annotation instance can be constructed by the system by combining the core and customized attributes, which can then be transferred to the server and stored in the backend databases.

6. Metadata resource classification

To support the classification interface described in Section 4.2, a flexible classification scheme is required. Most existing DL classification mechanisms assume a single type of resource and a single classification taxonomy, which make them unsuitable for heterogenous Web resources. Moreover, these mechanisms usually require the classification process to be done manually, often at the time the resource metadata are created.

To overcome these shortcomings, we propose a new classification scheme (Lim, Liu, & Goh, 2002) that is more suitable for digital library systems designed for distributed Web resources, such as G-Portal. The core of the proposed scheme is the *classification schema*, which consists of the classification rules that produce the categories and a taxonomy definition that determines how the categories are organized.

Each classification schema can be expressed in two forms, in a special language or its corresponding XML representation. The syntax of the language is similar to SQL and is more appropriate for humans to write; while the XML representation is appropriate for machine interpretation and interchange. A schema compiler has been implemented to convert the schema from the language version to its XML counterpart. In the following discussion, we will use the language to illustrate the features of the classification schema.

Currently, we allow only one metadata resource schema to be used in a classification schema, but more than one classification schema can be created for the same metadata resource schema. By having different classification schemas for the same resource schema, different taxonomies can be constructed for the same set of metadata resources. This achieves flexibility that is not found in existing DLs where resources are usually associated with a single taxonomy.

A classification schema consists of a *schema declaration*, one or more *classification rule definition*, and an optional *taxonomy definition* sections as shown in Fig. 11.

6.1. Classification rule

Consider the classification schema example shown in Fig. 12. The define schema statement describes the schema by providing the schema name (i.e., Schema1) and the resource schema (i.e., ExamQuestion.xsd). Two classification rules are defined, namely ByYear and ByTopicPhysical. Each rule specifies: (1) the

```
// classification schema declaration
define schema <classification schema name> on <resource schema>


// classification rule declaration
(define rule <rule name>
   classify by <resource element>
   [grouping ( ('{'(<list of values>|<integer range>)'}'|other) under <category name>)*]
   [where <selection condition>])+


// taxonomy definition
[define taxonomy
   (grouping '{'<list of category names>'}' under <category name>)+]
```

Fig. 11. Classification schema definition.

```
define schema Schema1 on ExamQuestion.xsd;
define rule ByYear
    classify by /Resource/Content/Paper/Year;
define rule ByTopicPhysical
    classify by /Resource/Content/Subject/Topic
    grouping {'Vegetation types', 'Human modification of natural vegetation'}
                under 'Natural vegetation',
            {'Coastal processes and resulting landforms', 'Coasts and human activities'}
                under 'Coasts',
            {'Development of a river system', 'River processes and resulting landforms',
                'Rivers and human activities'}  under 'Rivers',
            others under OtherTopics
    where /Resource/Content/Subject/Part = 'Physical Geography';
```

Fig. 12. A classification schema for exam question.

element used for classification; (2) the category names; and (3) the range of values to be applied on the element for each category.

In the ByYear rule, question resources are classified by their Year elements using the classify by clause. The elements or attributes used for classification are specified using XPath syntax. [5] The rule does not explicitly specify the category names and their value ranges. In this case, one category will be created

---

```
define schema Schema2 on ExamQuestion.xsd;
define rule ByNumber
      classify by /ExamQuestion/Content/Paper/Number
      grouping {1 ... 2} under 'Section1',
               {3 ... 5} under 'Section2',
               {6 ... 10} under 'Section3';
```

Fig. 13. Another classification schema for exam question.

```
define schema Schema3 on ExamQuestion.xsd;
define rule ByDifficulty
      classify by /ExamQuestion/Content/Description
      calling classifier1 under 'Easy',
              classifier2 under 'Intermediate',
              classifier3 under 'Difficult';
```

Fig. 14. Classification schema using classifiers.

for each Year value which is also used as the category name. This will result in 10 categories (for the past 10 years), since we are using a 10-year collection of questions.

In the ByTopicPhysical rule, question resources are classified by their Topic elements. This rule however defines four categories, namely Vegetation, Coasts, Rivers, and OtherTopics. [6] Each category is assigned a set of element values using the grouping clause. With this rule, question resources with topic matching either 'Vegetation types' or 'Human modification of natural vegetation' will be assigned to the Natural vegetation category. The other categories are similarly defined. Note that the others keyword allows us to specify a catch-all range of values. This keyword can only be used to define the last category in the grouping clause. The where clause also ensures that only questions of the first part of the syllabus (on physical geography) are classified by the ByTopic rule. Since no hierarchy is defined in the schema, the schema will produce a flat set of categories defined above.

As shown in Fig. 12, it is possible for a metadata resource to be classified into more than one category, e.g. 2002 and Physical categories, by using multiple rules. This can also occur within the same classification rule when categories are defined on elements that have multiple values, when different categories share some common element values, or when value ranges defined in the grouping clause overlap. An example of defining value range (without overlap) is shown in the ByNumber rule which classifies question resources into 3 categories (sections) based on the value of the Number element (Fig. 13).

So far, our discussion on classification rules focuses only on criteria that can be directly specified. In cases where the classification criteria require complex evaluation, such as applying text classification methods on textual elements, it is necessary to invoke external classifier programs. To cater to these situations, our classification language provides a calling clause to specify such invocations as shown in Fig. 14, where classifier1, classifier2 and classifier3 are executable classifier programs located within the DL. When evaluating the rules, the classification engine invokes and feeds the value of the actual resource's Description element to each of the classifiers and obtains a Boolean result to decide if the metadata resource belongs to

---

[6] This example has been simplified due to space constraints. The actual rule contains seven categories, including OtherTopics.

each category. Note that more than one classifier may return true for a resource, in which case the resource is classified into multiple categories.

6.2. Taxonomy definition

The categories defined by the previous classification rules only exist in a flat category space unless we further organize them into a hierarchy using the classification language. In a classification schema, taxonomy definition is optional but is usually desirable.

We use the Schema1 (see Fig. 12) as an example. Suppose we would like to group the topics resulting from the ByTopicPhysical rule and other rules in the same schema into themes, one possible taxonomy definition, as shown in Fig. 15, can be added to the previous classification schema. The definition taxonomy is contained within the define taxonomy statement. A set of grouping clauses essentially specify the parent–child relationship between a parent category and its child categories, by grouping subcategories into higher-level categories. For example, the base categories, Population, Settlements, Agriculture are grouped under a higher-level category known as Human. Given a classification rule, the resources not categorized by explicit element values are denoted by the others label. To distinguish between others labels of different classification rules, we disambiguate them using the rule name as their prefixes. In Fig. 15, assuming that we have another classification rule ByTopicHuman in the same schema, all categories that result from these two rules and are not mentioned in the previous two groupings will be grouped under the OtherThemes category. Note that categories without parents will be placed under the root category named using the name of the classification schema.

In this particular example, Physical, Human and OtherThemes are grouped under the Schema1 category, which is the root node of the category hierarchy. A graphical view of part of the taxonomy is shown in Fig. 16. It should be pointed out that if there are "unexpected" categories that are not captured by all the grouping clauses, they will be placed under the root category by default. "Unexpected categories" may be produced if we define a rule based on element values, such as the ByYear rule in Fig. 12.

```
define taxonomy
      grouping {'Natural vegetation', 'Coasts', 'Rivers'} under 'Physical',
      grouping {'Population', 'Settlements', 'Agriculture'} under 'Human',
      grouping (ByTopicPhysical.others, ByTopicHuman.others} under 'OtherThemes';
```

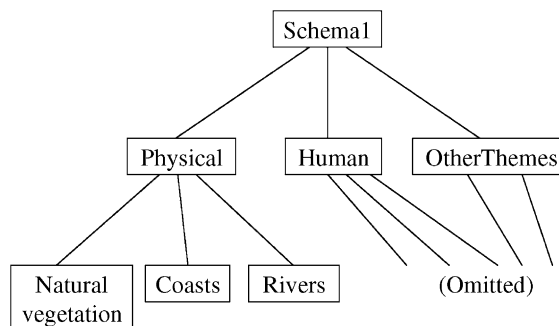Fig. 15. A taxonomy definition for exam questions.

Fig. 16. A graphical view of the taxonomy.

6.3. Classification engine

Given a classification schema expressed in the classification language, G-Portal provides a *classification schema compiler* to generate the XML version of the classification schema that can be shared over the Web and consumed by other applications.

The classification schema in XML will then be read by the *classification engine* which evaluates the classification rules on a set of metadata resources, generates the specified categories, and constructs a category hierarchy (taxonomy). Note that the actual classification of resources and the construction of category hierarchies is fully automated. The classification should take place prior to using the classification-based user interface. The resultant taxonomy and category assignment information will be stored in a category assignment file and the resources will then be grouped into folders and sub-folders corresponding to the category tree structure. This information can later be used by G-Portal's classification interface to display the classified resources and answer user queries. The category assignment file will have to be updated when there are changes to the resources attributes or the classification schema.

7. Conclusions and future work

In the G-Portal project, a digital library of geospatial and georeferenced resources is developed and serves as a unique Web portal capable of presenting information both spatially in a map-based interface and non-spatially in a hierarchical classification interface. G-Portal has been designed to support geography education use cases. With G-Portal, users can not only view these resources but also contribute geospatial and georeferenced resources under different projects using pre-defined resource schemas. By allowing resources to be shared across projects, G-Portal facilitates sharing of resources. In addition to visualization and query formulation, G-Portal also supports flexible classification schemes making it easier to adapt to different use cases for learning and research. In this paper, we described the system architecture of G-Portal and its modules. The design of its resource model, visualization, classification, and annotation capabilities are also discussed in detail.

7.1. Implementation

The implementation of G-Portal has almost completed. G-Portal's client-side modules have been implemented as a Java applet as this allows users to access the portal through Web browsers. 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. The implementation of the map-based interface employs an open source software package called GeoTools [7] which provides a set of Application Programming Interfaces that encapsulate the common functionalities required by geographical-based systems. The modules for manipulating projects, layers, resources and schemas are implemented in the Java applet as well.

The server is also implemented in Java to facilitate communication with the client. The server serves as a data broker between the client and the two databases Tamino XML Server and Informix Database. The former is used to store the non-spatial elements of G-Portal resources while the latter is used to store the spatial locations of resources and other relational information.

As of the writing of this paper, we have imported a simple dataset of general geographical resources, including country borders, rivers, lakes and cities. The metadata from DLESE has been harvested. We have also collected data from one of the FGDC clearinghouse server (Federal Geographic Data Committee

---

[7] http://www.geotools.org/

(FGDC), 2003). The Singapore GCE 'O' level geography exam questions are being populated into the G-Portal system as well (Chua et al., 2002).

7.2. Future work

As part of our future research activities, we will continue the implementation of G-Portal in the following research areas:

- *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. We are now conducting experiments on the use of G-Portal to support use cases of students preparing for the nation-wide geography examination in Singapore. A full scale evaluation of the pedagogical aspect of G-Portal is also in plan and this will involve both the geography teacher trainees in our institution and students studying geography in some high school.
- *Enhancements of Resource Schema Design:* The existing G-Portal resource schema has been designed to represent metadata resources with much flexibility. In the long term, when the resources are to be shared, across applications, we need to improve the resource schema further to facilitate interoperability. In particular, to give well-defined semantics to the metadata resource elements, namespace will have to be incorporated. We are also looking the definition of different user roles in the creation and maintenance of metadata resources.
- *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. This will allow G-Portal users to gather knowledge quickly from the Web.
- *Integration with e-learning systems:* To completely tap the potential of digital libraries, it is necessary to integrate digital library systems like G-Portal with e-learning systems. The merger of e-learning and digital libraries therefore needs to be carefully investigated. In particular, we are looking into how the e-learning metadata can be integrated with the G-Portal's resource schemas, and how the e-learning and G-Portal's functions can blend together to enhance the students' learning experience.
- *Automated/Semi-automated Metadata Resource Gathering:* At present, the resources hosted by G-Portal include high school geography examination questions and answers, DLESE resources, USA census, country, river and lake resources. Most of these metadata resources have been imported from external sources through specially written script programs. As part of its resource management and annotation modules, G-Portal also allows project administrators and users to contribute individual resources. These two approaches are not scalable since they require too much human involvement in script programming, source selection, and resource definition. Ideally, we would like to automate the metadata resource gathering process as much as possible using machine learning techniques. This will constitute an important part of our future research work.

References

ADN. (April 2003). ADEPT/DLESE/NASA (ADN) Joint Metadata Framework, http://piru.alexandria.ucsb.edu/metadata/.

Chua, L.-H., Goh, D. H.-L., Lim, E.-P., Liu, Z., & Ang, R. P.-H. (July 14–18, 2002). A digital library for geography examination resources. In *Proceedings of the second ACM+IEEE joint conference on digital libraries (JCDL 2002)*. Portland, OR, USA, pp. 115–116.

Coleman, A., Smith, T., Buchel, O., & Mayer, R. (2001). Learning spaces in digital libraries. In *Proceedings of fifth European conference on research and advanced technology for digital libraries (ECDL 2001)* (pp. 251–262). Berlin, Darmstadt, Germany: Springer.

Columbia Earthscape. (2002). Available from http://www.earthscape.org.

Federal Geographic Data Committee (FGDC). (2003). FGDC Clearing House, http://www.fgdc.gov/clearinghouse/clearinghouse.html.

Fowler, M., & Scott, K. (1999). *UML distilled: A brief guide to the standard object modeling language*. Addison-Wesley.

Goh, D., & Leggett, J. (2000). Patron augmented digital libraries. In *Proceedings of the fifth ACM conference on digital libraries (DL'00)* (pp. 153–163). San Antonio, TX, USA: ACM Press.

Kahan, J., & Koivunen, M.-R. (May 1–5, 2001). Annotea: An open RDF infrastructure for shared web annotations. In *Proceedings of the tenth international world wide web conference (WWW 10)*. Hong Kong, China, pp. 623–632. Available from: http://citeseer.nj.nec.com/kahanolannotea.html.

Lim, E.-P., Liu, Z., & Goh, D. H.-L. (July 8–12, 2002). A flexible classification scheme for metadata resources. In *Proceedings of digital library—IT opportunities and challenges in the new millennium (DLOC 2002)*. Beijing, China.

Liu, Z., Lim, E.-P., & Goh, D. H.-L. (December 11–14, 2002). Resource annotation framework in a georeferenced and geospatial digital library. In *Proceedings of the 5th international conference on Asian digital libraries (ICADL 2002)*. Singapore, pp. 287–398.

Liu, Z., Lim, E.-P, Ng, W.-K., & Goh, D. H.-L. (June 2003). On querying geospatial and georeferenced metadata resources in G-Portal. In *Proceedings of the third ACM+IEEE joint conference on digital libraries (JCDL 2003)*. Houston, TX, USA, pp. 245–255.

Orendorf, J., & Kacmar, C. (1996). A spatial approach to organizing and locating digital libraries and their content. In *Proceedings of the first ACM conference on digital libraries (DL'96)* (pp. 83–89). Bethesda, MD, USA: ACM Press.

Proulx, M.-J., Bédard, Y., Létourneau, F., & Martel, C. (December 1996). GEO-REP: A WWW customizable georeferenced digital library for spatial data. D-Lib Magazine.

Seber, D., Sandvol, S., Brindisi, C., & Barazangi, M. (November 2001). Building the digital earth. In *Proceedings of the geological society of America annual meeting*. Boston, MA.

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

Smith, T., Janee, G., Frew, J., & Coleman, A. (June 2001). The Alexandria digital earth prototype system. In *Proceedings of the first ACM+IEEE joint conference on digital libraries (JCDL 2001)*. Roanoke, VA, USA, pp. 118–119.

Sumner, T., & Dawe, M. (June 2001). Looking at digital library usability from a reuse perspective. In *Proceedings of the first ACM+IEEE joint conference on digital libraries (JCDL 2001)*. Roanoke, VA, USA, pp. 416–425.

Takahashi, J., Kushida, T., Hong, J., Sugita, S., Kurita, Y., Rieger, R., Martin, W., Gay, G., Reeve, J., & Loverance, R. (1998). Global digital museum: Multimedia information access and creation on the internet. In *Proceedings of the third ACM conference on digital libraries (DL'98)* (pp. 244–253). Pittsburgh, PA, USA: ACM Press.

Theng, Y.-L., Goh, D. H.-L., Lim, E.-P., Liu, Z., Pang, N. L.-S., Wong, P. B.-B., & Chua, L.-H. (December 11–14, 2002). Intergenerational partnerships in the design of a digital library of geography examination resources. In *Proceedings of the 5th international conference on Asian digital libraries (ICADL 2002)*. Singapore, pp. 427–439.

Xerox, C. M., 2000. The future of annotation in a digital (paper) world. In Harum, Twidale (Eds.), *Successes and failures of digital libraries*. Urbana-Champaign: University of Illinois, pp. 97–117. Available from http://citeseer.nj.nec.com/451771.html.