This document is downloaded from DR-NTU (https://dr.ntu.edu.sg) Nanyang Technological University, Singapore.

Resource annotation framework in a georeferenced and geospatial digital library

Lim, Ee Peng; Goh, Dion Hoe-Lian; Liu, Zehua

2002

Liu, Z., Lim, E. P., & Goh, H. L. (2002). Resource annotation framework in a georeferenced and geospatial digital library. Conference on Asian Digital Libraries (5th:2002:Singapore)

https://hdl.handle.net/10356/91116

The original publication is available at www.springerlink.com

Downloaded on 29 Mar 2024 22:01:35 SGT

Liu, Z.H., Lim E.P., and Goh, D.H. (2002) Resource annotation framework in a georeferenced and geospatial digital library. *Proceedings of the 5th International Conference on Asian Digital Libraries*, (December 11-14, Singapore), *Lecture Notes in Computer Science* 2555, 287-298.

Resource Annotation Framework in a Georeferenced and Geospatial Digital Library*

Zehua Liu¹, Ee-Peng Lim¹, and Dion Hoe-Lian Goh²

Centre for Advanced Information Systems, School of Computer Engineering Nanyang Technological University, Singapore, 639798, SINGAPORE {aszhliu, aseplim}@ntu.edu.sg
School of Communication and Information
Nanyang Technological University, Singapore, 639798, SINGAPORE
ashlgoh@ntu.edu.sg

Abstract. G-Portal is a georeferenced and geospatial digital library that aims to identify, classify and organize geospatial and georeferenced resources on the web and to provide digital library services for these resources. Annotation service is supported in G-Portal to enable users to contribute content to the digital library. In this paper, we present a resource annotation framework for georeferenced and geospatial digital libraries and discuss its application in G-Portal. The framework is flexible for managing annotations of heterogeneous web resources. It allows users to contribute not only the annotation content but also the schema of the annotations. Meanwhile, other digital library services, such as visualization and classification, can be provided over the annotations since they are treated as first class objects. This paper mainly focuses on the design of the resource annotation framework.

1 Introduction

There is a large amount of geospatial resources available on the World Wide Web. While these are valuable information for educational and research purposes, there have not been many georeferenced and geospatial digital libraries (DLs) developed for such public domain resources. In the DLESE project, metadata of earth related web sites/pages are contributed by user communities to establish a digital library consisting of quality education and resource materials [9].

To create a better learning experience and to allow sharing of users knowledge about resources, it is often desirable to allow users to annotate the resources with their own knowledge, store these annotations in the DLs and make them available to other users. Nevertheless, in most existing georeferenced and geospatial DLs [8, 2, 9], users cannot easily place comments on geospatial and georeferenced resources or share their knowledge about these resources due to the lack of direct support for annotation services in these systems. Users contribution of contents, including annotations, are usually not permitted or only permitted with certain

 $^{^{\}star}$ This work is partially supported by the SingAREN 21 research grant M48020004.

pre-defined format in most DLs. This greatly reduces the power of knowledge sharing.

Supporting annotations on web-based geospatial and georeferenced resources poses several challenges to the design of an annotation framework for geospatial and georeferenced digital libraries. As web resources are distributed and heterogeneous, it is unrealistic to impose a single format or even a pre-defined set of formats for all annotations. Users should be allowed to create annotations on different types of resources and on new types of resources that are unknown when the annotation framework was designed.

Besides supporting annotation on resources, digital library services must be provided to enable meaningful uses of the contributed annotations. For digital libraries of geospatial resources, annotation support should also be able to take geospatial attributes into consideration when offering services such as query and visualization.

In this paper, we propose a resource annotation framework for enabling contribution and sharing of user knowledge on geospatial and georeferenced resources. This framework is developed based on G-Portal [5], a georeferenced and geospatial digital library. In G-Portal, resource annotation is provided as one of the digital library services. The contributed annotations are treated as first class objects, i.e. they are considered as resources in the DL as well. Other digital library services can be provided over these annotations, by treating them as normal resources.

In G-Portal, each annotation is "typed" by associating an annotation schema with it. Annotation schemas are extensions of the basic resource schema. Apart from allowing users to add annotations to the DL, the framework also allows users to define customized annotation schemas and create annotations based on these new schemas. This gives the annotation framework the flexibility in dealing with heterogeneous resources.

1.1 The G-Portal Project

G-Portal [5] is an ongoing digital library project at the Centre for Advanced Information Systems in Nanyang Technological University. The aims of the project include the identification, classification and organization of geospatial and georeferenced content on the Web, and the provision of digital library services such as classification and visualization. In addition, authorized users may also contribute metadata resources to G-Portal, making it a common environment for knowledge sharing. In G-Portal, metadata resources are structured descriptions about Web content and annotations.

G-Portal also provides a platform for building applications that use geospatial and georeferenced content. This is achieved through **projects** that represent collections of metadata resources gathered for specific purposes or applications. Resources within projects are further organized into **layers** which allow finer grained logical organization of metadata resources.

Metadata resources within a project are visualized using either a map-based interface or a classification interface. The map-based interface displays resources

with spatial attributes. For example, metadata resources of countries, rivers and mountains and their associated content (such as identifying a particular climatic region on a map) can be shown in the map-based interface for easy viewing. Navigation tools such as zoom and pan are provided for users to browse the map. In addition, layers in a project may be shown or hidden, changing the visibility of the associated resources.

Resources with or without spatial attributes can also be displayed within the classification interface that categorizes and presents resources using classification schemas. Examples of such resources include general information about climate ("Why are land and sea breezes a feature of many coastal regions") and population ("What problems does the growth of squatter settlements create for large urban areas?")

The map and classification interfaces are synchronized so that when a resource on one interface is accessed, related resources on the other interface are highlighted as well. For example, if a user selects a resource in a region on the map interface, the classification interface will highlight related resources, which may appear under different categories. In both the map and classification interfaces, the full content of any selected metadata resources can also be viewed.

1.2 Paper Outline

The remaining sections of this paper are organized as follows. In Section 2, we describe the design of resource annotation. In Section 3, we present the annotation framework in the context of G-Portal. In Section 4, we compare our work with some of the related work. Finally, some concluding remarks are given in Section 5.

2 Design of Resource Annotation

In this section, we propose a new resource annotation framework consisting of a schema approach to represent annotations and their relationships with the annotated resources. A detailed description of the annotation and its schema, with examples is presented.

2.1 Representation of Annotations

In G-Portal, we distinguish the resources maintained by the DL system from those pre-existing web resources located at the public domain web sites. The former are referred as the *metadata resources*, while the latter are known as the *raw resources*. Metadata resources are the more structured versions of their raw counterparts and are contributed by users knowledgeable about the associated web sites or pages. To allow a more flexible use of metadata resources, we also allow metadata resources to be created without associating them with web resources. For the rest of our discussions, we shall use the terms "resource" and

```
<!-- BasicResource.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</pre>
             <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 Resources

"metadata resource" interchangeably unless in situations where explicit terms are required.

EXtensible Markup Language (XML) [3] has been chosen to represent resources in G-Portal. This facilitates sharing and publication of resources for other systems to use. The XML representation also supports easy transformation of resources from one format to another. Each resource must be created using some resource schema defined using XML Schema [7]. All resource schemas are derived from a base resource schema as shown in Figure 1. The base resource schema includes the common attributes of a resource such as identifier, location, and access control. Each derived resource schema can include other attributes relevant to the kind of resources covered by the schema. Each resource is assigned a unique identifier within the G-Portal system. The location attribute registers the geospatial properties (i.e., shape and location) of the resource. The access control attribute keeps the ownership and security information about the resource.

In order to treat annotations as first-class objects, we can either treat them as a new class of objects or as a special type of resources. In our proposed framework, the latter representation option is adopted. In other words, we also introduce a schema for describing a set of annotations. This will be elaborated in Section 2.2. Other than the attributes of basic resource, a new attribute (known as the AnnotatedResources element) to identify the annotated resources is added to the annotation resource. Making annotation a subtype of resource allows us to treat annotations as resources. This also makes it possible for annotations to be annotated, just like ordinary resources. In this way, the DL services designed for normal resources can still be applicable to annotation resources. From the implementation point of view, the DL system will not be unnecessarily complex because of the introduction of such resource annotation framework, since there is still only one type of first-class objects to deal with.

```
<!-- 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>
            </r></rad:extension>
        </xsd:complexContent>
   </xsd:complexType>
    <xsd:complexType name="AnnotatedResourcesType">
            <xsd:element name="Resource" type="xsd:string" max0ccurs="unbounded"/>
       </xsd:sequence>
   </xsd:complexType>
</xsd:schema>
```

Fig. 2. The Base Schema for Annotations

Another side effect of an annotation being a resource is that annotations may have geospatial attributes as other resources do. The geospatial attributes will usually be derived from the resources being annotated, if they have geospatial attributes. An annotation may also has its own geospatial attribute value directly assigned by the annotator.

Unlike most existing annotation frameworks where annotations can only be created for individual resources, our framework allows annotating multiple resources with a single annotation. For example, in a map that displays South-East Asia countries, a comment about the historical relationship between Malaysia and Singapore can be made by annotating the resources representing the two countries. In the traditional annotation systems, this can only be achieved by annotating one country and mentioning the name of the other country inside the annotation itself. For instance, an annotation is created for the resource corresponding to Malaysia and this annotation carries a reference to the Singapore resource as its attribute. While this approach might sometimes achieve the same effect as annotating on both resources, it is not intuitive because it requires one of the annotated resources to be used as the anchor. To manipulate this annotation, one has to take an extra step to find out all other annotated resources from this anchor. Hence, our proposed framework has taken the option to allow annotation based on multiple resources and provides a standard way of representing the annotated resources as a list under the AnnotatedResources element.

2.2 Annotation Schema

As shown in Figure 2, the basic annotation schema extends the basic resource schema by appending a new AnnotatedResources element to keep the resource ids of the annotated resources. Other important attributes of a resource, such as id and geospatial attribute, are inherited from the basic resource schema.

Fig. 3. A Simple Resource about Singapore

Fig. 4. A Resource Schema for Country

The definition of the content of 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.

2.3 Annotation Example

This section gives an example of annotation represented in XML. An example of a resource about Singapore is shown in Figure 3 using a country resource schema given in Figure 4. Note that basic attributes of a resource include id, name, location, creator, source, and the actual content. Details of some attributes are omitted because they are not relevant to the annotation aspect. Resources about other countries, such as China and Malaysia, are represented similarly.

Before one can annotate the countries, a country annotation schema is first derived from the basic annotation schema shown in Figure 2. The derived annotation schema is shown in Figure 5. The new annotation schema can be derived by altering the definition of the "Content" node in the basic annotation schema.

An annotation about Singapore and Malaysia is shown Figure 6. Detailed definition of the individual elements is omitted due to space constraint. Comparing the format to that of the resource, one additional element "AnnotatedResources" is inserted, which is used to indicate the id of resources to be annotated. In this example, we are annotating the Singapore and Malaysia (resource id is "Country_60") country resources. The "Content" element specifies the details of the annotation. Here, only a simple comment is used to describe the relationship of physical locations between the two countries. The structure of the "Content" element depends on the annotation schema and is not fixed.

Fig. 5. An Country Annotation Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<Resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"</pre>
         xsi:noNamespaceSchemaLocation="Annotation.xsd">
    <ID>Annotation_01</ID>
    <ResourceName>
        <Name>Annotation about Singapore and Malaysia
    </ResourceName>
    <Location Type="Geometry"> <Geometry> ... </Geometry> </Location>
    <Creator> ... </Creator>
    <Source> ... </Source>
    <Content>
        <Comment>Singapore is located beside Malaysia</Comment>
    </Content>
    <AnnotatedResources>
        <Resource>Country_65</Resource>
        <Resource>Country_60</Resource>
    </AnnotatedResources>
</Resource>
```

Fig. 6. A Simple Annotation about Singapore and Malaysia

3 Annotations in G-Portal

This section describes how the resource annotation framework is applied in the G-Portal system. We will focus mainly on the functions of creating annotations, browsing annotations and classification of annotations.

3.1 Creating an Annotation

Registration of Annotation Schemas As annotations must be instantiated by some annotation schemas, the annotator must register his or her annotation schemas with G-Portal. The annotation schemas can be defined beforehand and made available as XML Schema files on the web. The registration of an annotation schema simply involves providing the URIs of these schema files to G-Portal.

Creating Annotation Layer in a Project Recall from Section 1.1 that a project contains a collection of resources and all these resources are organized into layers before grouped into the project. Annotations, as some kind of resource,

should also be grouped into some layer in the project in order for the G-Portal to visualize them, since the G-Portal only visualizes resources in the current project.

Each project has an owner. Before one can annotate resources within a project, the project owner must configure the project layers and their resources to be readable by the public users. The owner should also grant public users the permission to create new layers within the project. Such permission, while applicable to the new annotation layers only, will not affect the original layers constructed by the owners.

When the necessary access rights are granted, an annotator can first create a new layer under the current project. The new layer will have the annotator as the owner. Nevertheless, since the annotation layer belongs to the project, its existence will also be dependent on the project.

Bookmarking in G-Portal The next step to create an annotation is to indicate the resources to be annotated. This is accomplished by the *bookmarking* facility provided by the G-Portal. The bookmarking facility was initially developed to allow easy exploration of resources between the *map-based interface* and *classification interface* in G-Portal, where users can bookmark target resources in one interface and continue exploring among these resources in another interface.

Bookmark resources are created incrementally by repeatedly selecting resources and invoking the command "Add to Bookmark". Users can also ask G-Portal to select all bookmarked resources or remove some resources from the bookmark. Bookmarked resources are highlighted differently from those not bookmarked and the normal selected resources. They will not be affected by normal selection operation.

Bookmarks in the two interfaces are synchronized. Resources bookmarked in one interface will also be highlighted in the other, so that users can easily switch between the two interfaces when looking for certain resource. For example, the user can draw a rectangle to select resources in the map-based interface and bookmark them, and then go to the classification interface to explore the bookmarked resources one by one to pin down the resources that he or she is interested in.

To facilitate annotation, bookmarking is used to select the resources to be annotated. This is important especially when users want to annotate multiple resources, each locating in different place in the map.

Construction of an Annotation Once the resources to be annotated are bookmarked, a new annotation record can be constructed within the annotation layer by selecting the appropriate annotation schema to be used. This annotation schema can be selected from a list of annotation schemas registered by the annotator.

After the annotation schema is selected, G-Portal provides an interface, as shown in Figure 7 for the user to enter the content of the new annotation. Since

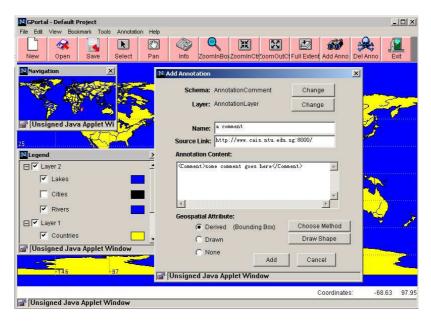


Fig. 7. The Add Annotation Dialog

the content of annotations can be defined freely by users, there is no fixed interface for users to enter the content. Moreover, the XML hierarchical format of annotation content makes it more difficult to provide a meaningful interface for content creation. In the current implementation of the G-Portal, we have decided to provide a simple text editor for users to enter XML contents. The user is allowed to enter any XML text conforming to the annotation schema. The user also provides values of other attributes such as Name and Link. The rest of attributes including ids of annotated resources and geospatial attribute will be automatically generated and combined with the XML content entered to form the complete annotation. Note that the ids of the annotated resources are automatically identified and inserted into the annotation as the "AnnotatedResources" attribute.

Geospatial Attribute of Annotations In G-Portal, annotations, just like resources, may contain geospatial attributes that describe their geographical representation. The annotator may directly provide the G-Portal with the geospatial information, if the annotation itself has such kind of information. The user provides the shape information by choosing the "Drawn" option in Figure 7 and drawing the shape on the map. For example, when commenting on the origin of Singapore population, the user may want to use the geographical area of the southern part of China as the geospatial property.

Alternatively, the user may ask the G-Portal to automatically derive this geospatial information by performing certain spatial operation on the geospatial attributes of the annotated resources. This is done by selecting the "Derived"

option in Figure 7 and clicks on the "Choose Method" button to select an appropriate derivation method. The selected method will be shown within brackets. One of such operations is to simply use the geospatial attributes of all the annotated resources. Another possible operation is to use the smallest rectangle that covers all resources, i.e. the bounding box. This is especially useful when annotating a group of resources locating near each other, such as the ASEAN (Association of South-East Asian Nations) countries.

3.2 Browsing Annotations

As mentioned earlier, the G-Portal provides two interfaces, map-based interface and classification interface, for presenting resources. Users can browse through annotations using these two interfaces, just like what they do with normal resources.

The map-based interface is more intuitive but only limited to annotations with geospatial attribute. Users can use the navigation tools provided by the G-Portal to navigate the map and to zoom in and zoom out till the target information is located. The classification interface shows all annotations with or without geospatial attribute. Annotations are classified into categories according to some classification schema [6]. The categorized annotations are then presented in the classification interface in the tree structure determined by the category hierarchy. Users browse through the classification interface in a similar way to looking for a file through the Windows Explorer.

To further assist the navigation of resource annotations, when an annotation is selected by the user, the annotated resources by that annotation will be automatically highlighted as well. In this way, the relationship between annotations and annotated resources becomes more obvious and can be identified by users more easily.

Resources that have been annotated will also be presented in a different highlighting style. When users see a resource being highlighted in a different style, they can request for all the annotations that have been associated with the resource. The result of the request is a list of annotations, which the user can go through one by one.

Users can view the content of annotation in the same way as that of normal resource. The current implementation of the G-Portal displays resource content in plain XML format, with indentation formatting. Users view the content of annotation by double-clicking on the annotation or selecting the Info Tool and single-clicking on the annotation.

3.3 Classification Services for Annotations

The classification interface of the G-Portal is capable of showing annotations and presenting them in a hierarchical interface. The classification interface is supported by the classification service provided by the G-Portal's backend system. The classification is based on rules that rely on values of certain attributes of the resources. The actual classification of annotations is performed by the

classification engine on the server. The classification produces a hierarchy of categories with annotations assigned to the categories. The result is presented in the classification interface when requested.

Apart from defining the classification rule based on standard attributes of resources, special attribute, especially the ids of the resources being annotated, can be used to provide useful classification for annotations. For example, based on the ids of the annotated resource, annotations about countries may be grouped together by countries and further by regions. This kind of classification would not be possible for normal resources without introducing extra attributes into the resource to indicate the name or id of the countries being annotated.

4 Related Work

Annotation has always been seen as an important digital library function that solicits user contributed knowledge which can be further shared among users [11].

In the context 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) [4]. Each annotation is represented as a set of metadata 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..). G-Portal is also similar to Annotea in providing some database servers to store the created annotations. On the other hand, G-Portal and Annotea has a few notable differences. Firstly, G-Portal's annotations are created over the metadata resources of the web objects instead of the web objects directly. Secondly, G-Portal provides an additional annotation element to accommodate the geospatial property of annotations. Thirdly, the content of annotations in Annotea is not structured (free text or HTML document) and the annotations can be associated with a single document and some portion of it. Finally, G-Portal provides both map and classification views on the annotations.

In the DLESE and ADEPT digital library projects, a common metadata framework known as ADN Framework has been proposed to standardize the representation of metadata resources including annotations [9, 8, 1]. Annotations are mainly defined for educational purposes. The metadata format of annotation consists of of mainly the contributor, creation date and description components. The description component of an annotation is essentially a free text comment on the corresponding annotated web object. G-Portal, on the other hand, has adopted a more flexible annotation schema structure which allows basic annotation schema to be extended with different elements that capture a wide variety of annotations. The ADN Framework is currently under some major changes to adopt XML Schema for describing metadata.

Wilensky, in the UC Berkeley Digital Library project, proposed a new *multivalent document* model that supports documents with behaviors that allows annotations to be added to documents and be manipulated with a variety of operations, e.g. copyediting [10]. This document model has been implemented

in a GIS Viewer browser that supports addition of annotations to a document containing geospatial and multimedia information. Nevertheless, the concept of annotation schemas and collection-level visualization and classification of annotations were not included in the work.

5 Conclusion

In this paper, we proposed a resource annotation framework in the context of a digital library for geospatial and georeferenced web resources. The framework is flexible to handle the heterogeneity of web resources and different annotation needs. Annotations created in the framework are treated as first-class objects such that existing digital library services, such as classification and visualization, can be reused for them. The framework has been built into the G-Portal system.

References

- ADEPT/DLESE. ADN joint metadata content model. http://www.dlese.org/metadata/.
- 2. H. Chen, B.R. Schatz, T.D. Ng, J.P. Martinez, A.J. Kirchhoff, and C. Lin. A parallel computing approach to creating engineering concept spaces for semantic retrieval: The Illinois Digital Library Initiative Project. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):771–782, August 1996.
- 3. eXtensible Markup Language. http://www.w3c.org/xml/.
- 4. Jose Kahan and Marja-Ritta Koivunen. Annotea: an open RDF infrastructure for shared web annotations. In *Proceedings of the Tenth International World Wide Web Conference (WWW 10)*, pages 623–632, Hong Kong, China, May 1-5 2001.
- Ee-Peng Lim, Dion Hoe-Lian Goh, Zehua Liu, Wee-Keong Ng, Christopher Soo-Guan Khoo, and Susan Ellen Higgins. G-portal: A map-based digital library for distributed geospatial and georeferenced resources. In *Proceedings of the Second ACM+IEEE Joint Conference on Digital Libraries (JCDL 2002)*, Portland, Oregon, USA, July 14-18 2002.
- Ee-Peng Lim, Zehua Liu, and Dion Hoe-Lian Goh. A flexible classification scheme for metadata resources. In Proceedings of Digital Library – IT Opportunites and Challenges in the New Millennium (DLOC 2002), Beijing, China, July 8-12 2002.
- 7. XML Schema. http://www.w3c.org/xmlschema/.
- 8. 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.
- 9. 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.
- 10. Robert Wilensky. Digital library resources as a basis for collaborative work. *Journal of the American Society of Information Science*, 51(3):228–245, 2000.
- 11. Catherine Marshall Xerox. The future of annotation in a digital (paper) world. In Harum and Twidale, editors, *Successes and Failures of Digital Libraries*, pages 97–117. Urbana-Champaign: University of Illinois, 2000.