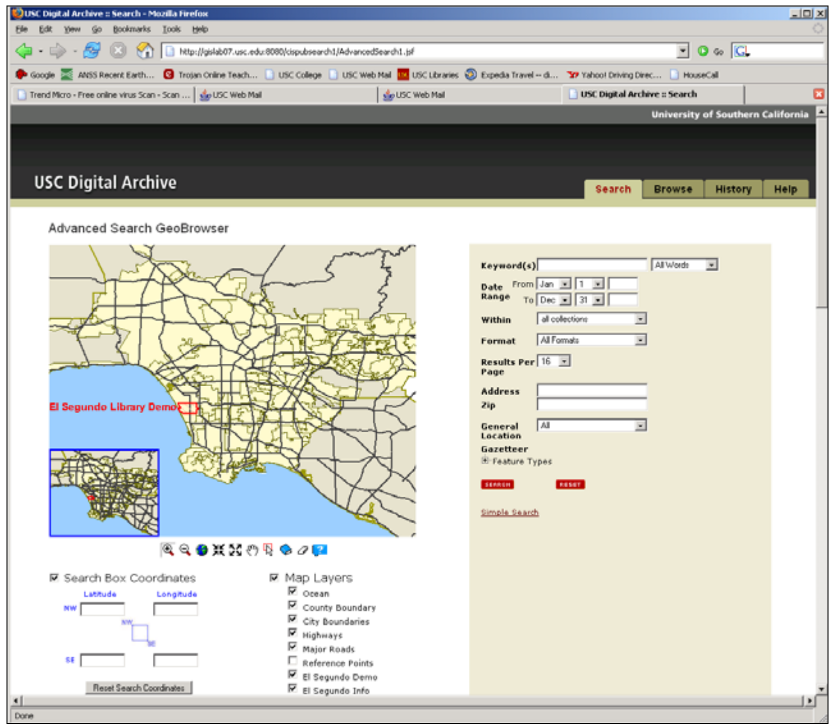
Prototype Geographic Search and Query Tools for the USC Digital Archive



Jennifer N. Swift
Parisa Ghaemi
Daniel W. Goldberg
John P. Wilson

University of Southern California GIS Research Laboratory
Technical Report No. 4

Prepared for: Digital Information Management
Collection Information Systems
Information Services Division
University of Southern California
Los Angeles, CA 90089

GIS research laboratory
Department of Geography
College of Letters, Arts, and Sciences
University of Southern California
Los Angeles, CA 90089-0255
www.usc.edu/dept/geography/gislab

Cover Photo:

Project application demo hosted on USC GIS Research Laboratory server:
http://gislab07.usc.edu:8080/cispubsearch1/AdvancedSearch1.jsf

Preferred Citation:

Swift, J., P. Ghaemi, D.W. Goldberg, and J.P. Wilson. 2005. *Prototype Geographic Search and Query Tools for the USC Digital Archive*. GIS Research Laboratory and Collection Information Systems (CIS), a subdivision of Information Service Division (ISD), University of Southern California Los Angeles, California.

Introduction

The GeoBrowser application was built for the USC Collection Information Systems (CIS) Digital Information Management (DIM) team, by the USC GIS Research Laboratory to provide the existing USC Digital Archive website with advanced spatial search capabilities. Three new spatial services were added to the existing website – a clickable map, an address geocoder, and a gazetteer. The GeoBrowser Advanced Search screen is the second main screen of the USC Digital Archive website (Figure 1). The user comes to this page by clicking the "Advanced Search" text link on the USC Digital Archive Simple Search screen.

The GeoBrowser Advanced Search screen is made up of the standard header, footer and grey background, and a clickable map. The other elements on the screen are the text-based Search tools, which include the address geocoder and gazetteer, and the text link to the Simple Search screen. These services give the user three new filters for guiding the identification and selection of content. After any searches or other information gathering, a user may return to either the Advanced Search or the Simple Search screen.

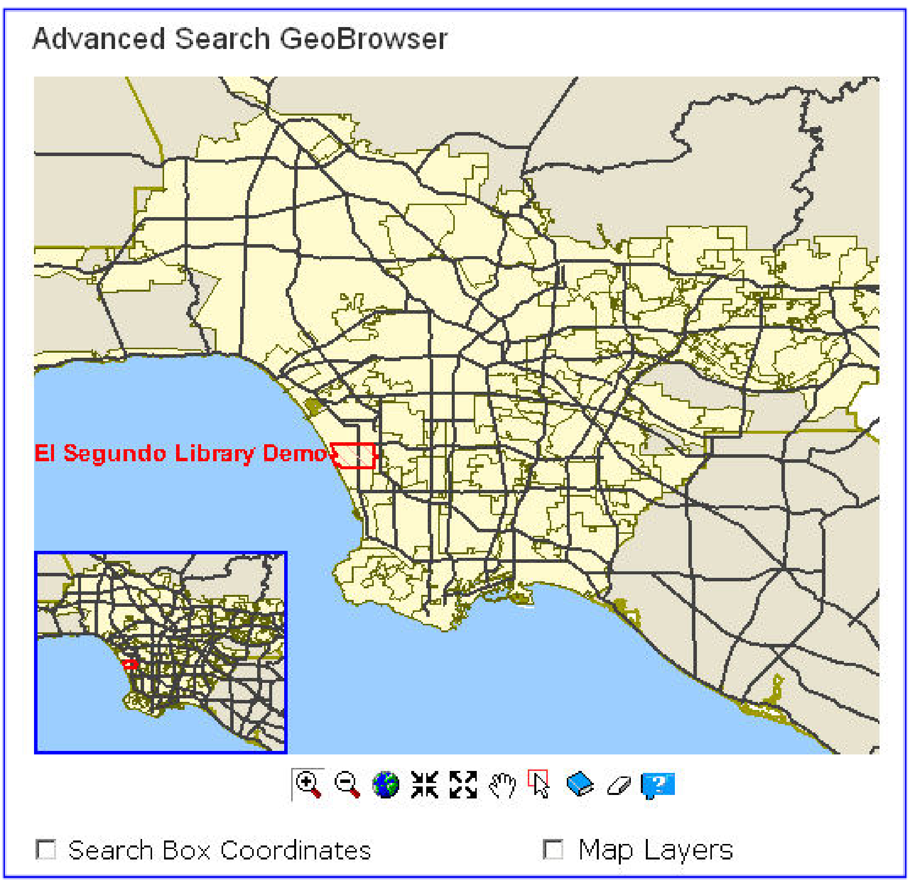


Figure 1. USC Digital Archive Advanced Search Geobrowser webpage.

This document provides an overview of the GeoBrowser Advanced Search demonstration application, for El Segundo, CA.

Geobrowser

The GeoBrowser consists of a clickable map that was custom-built for this project in Java, JSP and Java Server Faces (JSF) (Figure 2). The map application utilizes functions from ESRI's (Redlands, California) internet mapping software, ArcIMS.

Figure 2. Geobrowser clickable map, navigation and data retrieval tools. The red boundary delineates the demonstration area used for the current Gazetteer implementation.

Geobrowser map navigation tool buttons including Zoom In, Zoom Out, Zoom Global, Fixed Zoom In and Out and Pan were added to assist the user in finding locations or areas of interest on the map (Figure 2). The Zoom In and Out buttons allow a user to draw a rectangle on the map to zoom in or out to the desired area (Figure 3). The Fixed Zoom In and Out Buttons allow zooming using a fixed interval for each mouse click. The Zoom Global button returns the user to the original map view (i.e. scale and extent). The Pan Button can be used to click and drag the map in any direction.

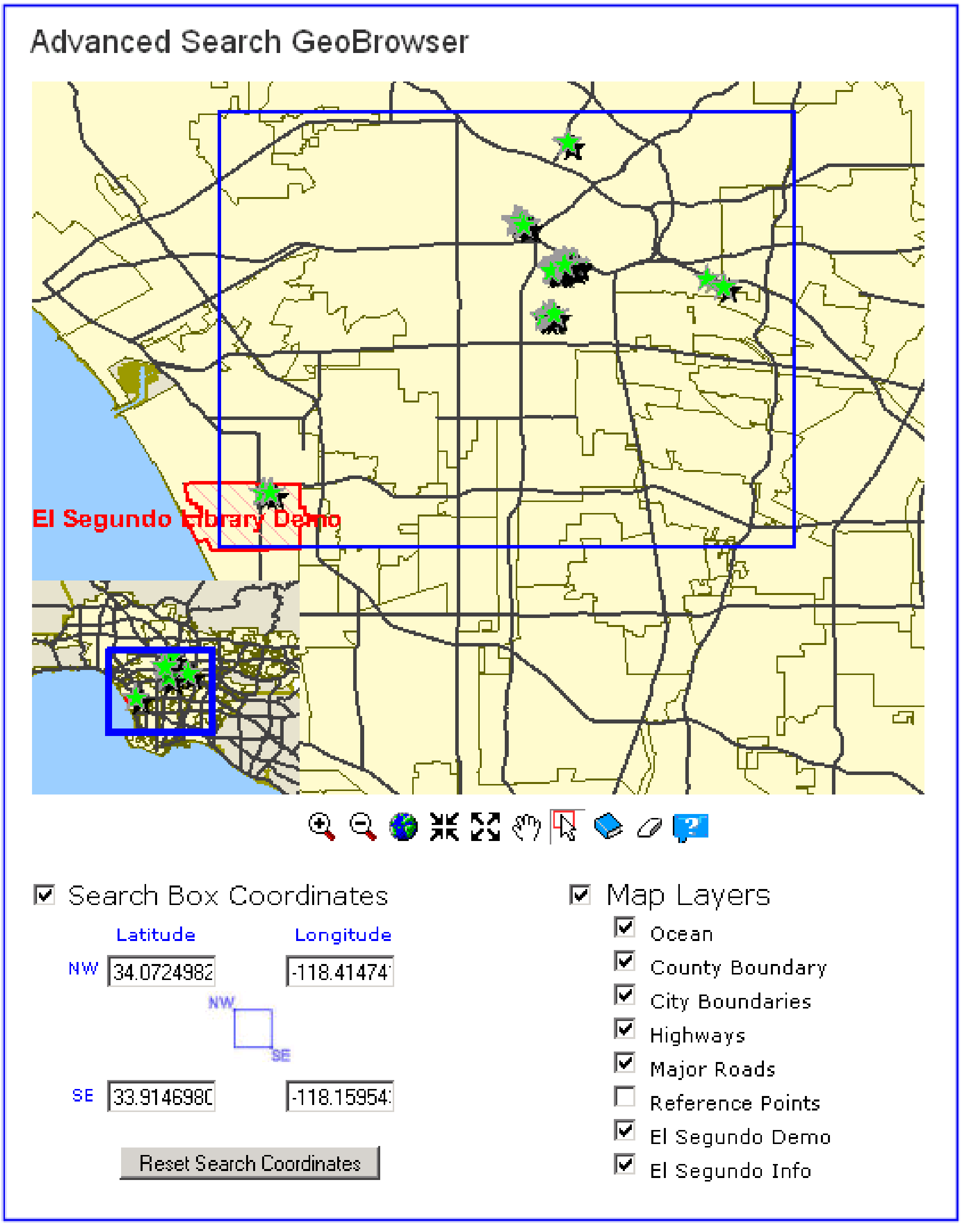


Figure 3. Blue box is drawn using the Search Box button. Selected objects (database records) can be plotted in area drawn on map using the Search Box button, and displayed

using the View Library Holdings button. The Search Box coordinates and visible map layers can be viewed beneath the map.
The primary GeoBrowser search tool is the Search Box button (Figures 2 and 3). The user can draw a box on the map using the “Search Box” button to retrieve all of the georeferenced records in the Digital Library database within the geographic boundaries of the box. The bounding coordinates of the SearchBox (latitude and longitude of NW and SE corners of a rectangle or box drawn on the map using the button) can be viewed in text boxes below the map (Figure 3). A user may use the View USC Digital Archive Holdings tool (button) to display the results of a Search Box search as points (graphic stars) on the map. Selections may appear in clusters, which may be better viewed by zooming in on the map. A user may also use the Clear Map View and Web Page tool (button) to clear the Search Box (blue rectangle) and Digital Archive holdings (points = stars) from the map view, and refresh the text-based search options (Figure 2). A Help tool (button) is also available to view GeoBrowser tool button descriptions.

Finally, Map Layers may be controlled through a list of check boxes available below the map (Figure 3). Layers can be turned on or off by clicking on the boxes adjacent to each layer name. This displays or hides the layers in the map view, respectively. The list of layers shown beneath the map consist of ESRI point, line and polygon shapefiles.

Geocoder

The geocoder was built independently as part of another USC research project to test the feasibility and scalability of new address matching procedures and algorithms developed by Bakshi et al. (2004; Goldberg et al. 2006). The main novelty of these algorithms is the consultation of multiple sources to discover the actual number of addresses (i.e. features) occurring on a street, their layout and orientation, and their actual geographic footprints, instead of using the address range associated with the georeferencing segment for linear interpolation as traditional geocoders do. This additional knowledge allows for locations along street segments with improved accuracy. These algorithms have been implemented into a scalable and reliable geocoding web service, which is integrated into the GeoBrowser application (Figure 4).

The geocoder can be used to search the USC Digital Library for selections without the requirement of the user knowing the spatial coordinates of their area of interest. A user who is unfamiliar with the geography of a region may not be able to locate and draw a region of interest on a map. Instead, a simple search based on address may be easier (Figure 4).

The user must enter both the street address and zip code for the geocoder application to work efficiently (Figure 4). An address and zip code search can be defined by entering any address in the United States, to search for selections within a given radius of the address. The radius is currently pre-defined in the application by the Digital Library staff. The zip code may narrow the geocoding search to a smaller area than a search on a street name. The search may be unsuccessful unless the user can also provide the zip code in those instances in which there are many streets with the same name. The search is executed by clicking on the red Search button (SEARCH) beneath the text-based Advanced Search options (bottom left side of Figure 5). All entries and selections in the text-based search tools can be cleared by clicking on the red "Reset" button (RESET).

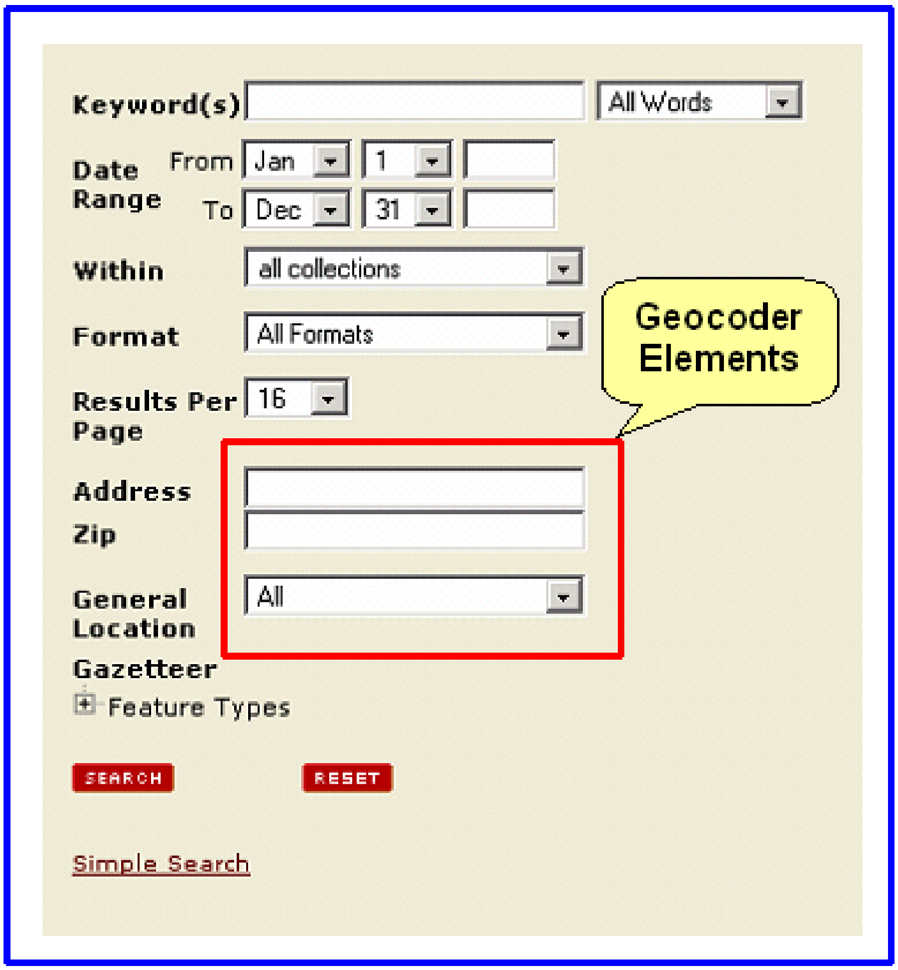


Figure 4. The Address and Zip text fields together comprise the Geocoder search tool.

Gazetteer

The gazetteer was built independently of this project as well (Goldberg et al., 2005a and 2005b). The current Gazetteer database and service is restricted to the city of El Segundo in Los Angeles, and was built to demonstrate how some newly developed web harvesting tools could be deployed to build gazetteer databases quickly and cheaply. The gazetteer provides a highly detailed location filter which the user can use to identify a specific location, such as a business, and retrieve all of the Digital Library selections within a specified geographic buffer around the chosen location (Figure 5). The buffer radius is currently pre-defined in the application by the Digital Library staff.

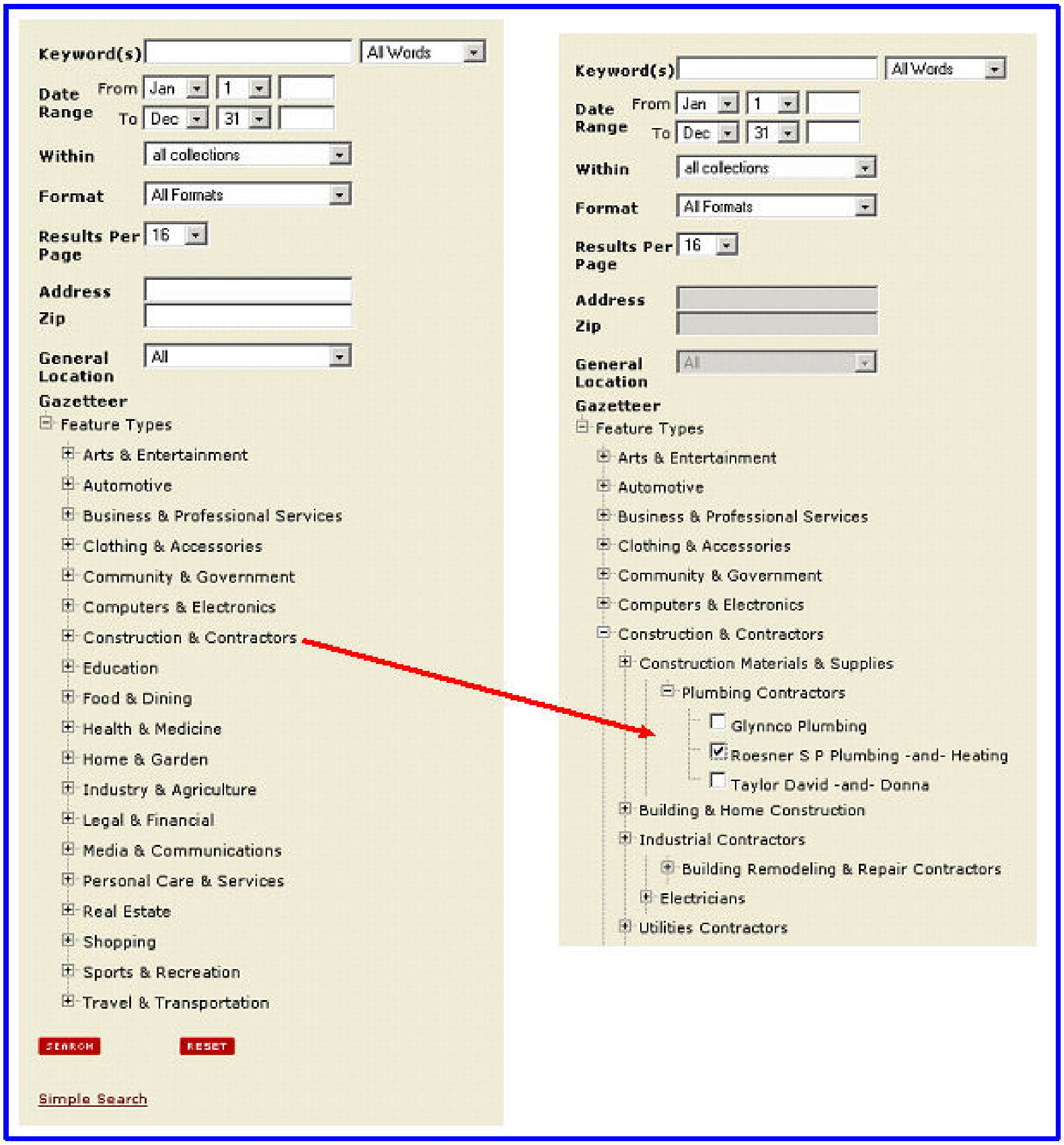


Figure 5. The Gazetteer search tool provides a hierarchical dropdown list of feature types. Selection of specific features such as businesses or landforms, support feature spatial searches.

The user clicks on the + box beneath gazetteer (Figure 5) to expand a hierarchical drop-down list of gazetteer feature types. Then a gazetteer-based search is performed by clicking any number of checkboxes adjacent to the feature type name, such as a construction material and supply business (right side of Figure 5). The search is executed by clicking on the red Search button (

) beneath the text-based Advanced Search options (bottom left side of Figure 5). All entries and selections in the text-based search tools can be cleared by clicking on the red “Reset” button (RESET).

The geographic “features” are described by three elements: type, name, and geographic footprint. Each of these feature elements is important to the user as well as the Digital Library staff. For instance, a user may be interested in locating library holdings in relation to a named geographic feature or type of feature, but not know the exact spatial coordinates of the feature of interest. The gazetteer can transform a named geographic feature into a spatial representation that is then used to search the database. Additionally, in the future the Digital Library staff could use the gazetteer as a tool to specify geographic footprints for library holdings quickly and cheaply. This is important because the geographic search and query tools discussed herein will work as planned only for Digital Archive holdings that incorporate geographic references (i.e. footprints).

Application Installation Notes

This application has been developed and tested with Tomcat 4.1.29. The Java Server Faces (JSF) API is used to render HTML pages, and the ESRI ArcIMS Java Connector API is used to communicate with an ESRI ArcIMS Image service. The application is also coded in JavaScript and CSS to provide the website with interactive GIS functionalities. Some additional guidelines for setting up the Geobrowser application are provided below.

System Requirements

Prior to setting up this application, perform the following tasks.

1. Have ArcIMS Server successfully running.
   Note: ArcIMS Java Connector jar file should have been added to the web application: CISDigitalArchive\WebContent\WEB-INF\lib folder.
2. Have a Servlet Engine successfully running.
   Note: This application has been developed and tested with Tomcat 4.1.29. The Tomcat version will be dependent on the ArcIMS version/installation.
3. Java Server Faces jar files and JSTL 1.0 jar files have been added to the web application: CISDigitalArchive\WebContent\WEB-INF\lib folder.

Setting Up The Web Application

Perform the following tasks to complete the Geobrowser application installation.

1. Extract the zip file provided. Verify that the extracted folder has the following structure:
   CISDigitalArchive
   |--ant

|--build
|--ExtLib
|--JavaSource
|--WebContent

2. The Apache Ant tool may be used to build and deploy the web application. First install the Apache Ant tool from the following location before proceeding any further: http://ant.apache.org.
3. Stop Tomcat Service.
4. Use the ant tool to build and deploy the web application:
   a) Open command prompt and cd to CISDigitalArchive/ant folder.
   b) Type: <drive>/<path to apache ant installation directory>/bin/ant build
   This will compile the java source code. Fix any code errors you receive while compiling.
   c) Type: <drive>/<path to apache ant installation directory>/bin/ant deploy.
   Verify that the CISDigitalArchive.war file is created in the CISDigitalArchive/build folder and also copied over to <tomcat_home>/webapps folder.
5. Create a web application in Servlet Engine:
   a) Copy CISDigitalArchive\WebContent folder to Tomcat_Home\webapps folder.
   b) Rename WebContent folder in Tomcat_Home\webapps directory to CISDigitalArchive.
6. Start Tomcat Service.
7. Test the web application by typing:
   Type the following URL address in the web browser:
   http://localhost:8080/CISDigitalArchive/index.jsp.
8. Verify that the tomecat_home\webapps\CISDigitalArchive\WEB-INF\src and tomecat_home\webapps\cispubsearch1\WEB-INF\lib\GISLab.jar have been added to CLASSPATH.

Conclusions

The GeoBrowser application was built to provide the USC Digital Archive website with advanced spatial search capabilities. This demonstration application concentrates on the city of El Segundo, CA. The spatial services include a clickable map, an address geocoder and a gazetteer. The map allows website users to easily define a geographic area of interest on a map, view available holdings within the selected area, and search the Digital Archive for information on those holdings. The ArcIMS map browser is coded in Java, JSF, JavaScript and CSS. The geocoder is written in C# and hosted as a web service. The geocoder provides another very popular kind of search tool. The user can enter any address and zip code in the United States to locate library holdings in the proximity of the address. The gazetteer provides a lengthy, detailed geographic “feature” type hierarchy. Gazetteer features are organized by type, name and geographic footprint, and this information is used as spatial input for searching the Digital Archive. All three of these spatial search tools, as well as the original Advanced Search page text-based search tools, can be used independently of one another to search for library holdings. Any combination search using the GeoBrowser and other text-based search options allows the user to search for library holdings in the Los Angeles County area.

Literature Cited

Bakshi, R., Knoblock, C., Thakkar, S., 2004. *Exploiting Online Sources to Accurately Geocode Addresses*, in Proceedings of the 12th International Symposium on Advances in Geographic Information Systems, ACM-GIS '04. http://www.isi.edu/info-agents/papers/bakshi04-acmgis.pdf.

Goldberg, D.W., Wilson, J.P. and Knoblock, C.A., 2005a. *Extracting geographic features from the internet to automatically build detailed regional gazetteers*. Unpublished paper presented at Second Annual ECAI Cultural Congress, Shanghai, China, 9-13 May.

Goldberg, D.W., Wilson, J.P. and Knoblock, C.A., 2005b. *Extracting geographic features from the internet to automatically build detailed regional gazetteers*. International Journal of Geographical Information Science (submitted 12/05).

Goldberg, D.W., Wilson, J.P. and Knoblock, C.A., 2006. *Beyond linear interpolation: Moving towards mediator-based geocoding using multiple sources*. Journal of the Urban and Regional Information Systems Association (submitted 1/06).