University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Papers in Natural Resources | Natural Resources, School of

11-1-2005

Description of a database support tool for retrospective georeferencing of natural history museum specimen collections

Rachel A. Simpson
University of Michigan, rsimpson2@unl.edu

Jeremy B. Young
University of Michigan

Follow this and additional works at: http://digitalcommons.unl.edu/natrespapers

Part of the Natural Resources and Conservation Commons

Simpson, Rachel A. and Young, Jeremy B., "Description of a database support tool for retrospective georeferencing of natural history museum specimen collections" (2005). *Papers in Natural Resources.* Paper 1.
http://digitalcommons.unl.edu/natrespapers/1

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Description of a database support tool for retrospective georeferencing of natural history museum specimen collections†

Rachel A. Simpson[1, 3] and Jeremy B. Young[2]

[1]Shapiro Science Library, University Library, The University of Michigan, Ann Arbor, MI 48109-1185
[2]Space Physics Research Laboratory, Department of Atmospheric, Oceanic and Space Sciences, College of Engineering, University of Michigan, Ann Arbor, Michigan 48109-2143
[3]Current address: School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68588-0517.
E-mail: rsimpson2@unl.edu

Introduction

The locations at which natural history museum specimens were collected ('localities') can be visualized using Geographic Information Systems (GIS) technology when explicit and standardized references to geospatial locations are available. Below we describe a database support tool which automates the association of localities with explicit geospatial coordinates for named places included in an electronic gazetteer. The tool furthermore facilitates simple manual review and correction of results.

Description of Database Support Tool

A locality consists of a string of text which may include one or more place names that also appear as entries in a gazetteer. The database support tool we developed links a set of specimen localities with entries from an electronic gazetteer. In our work we used the Geographic Names Information System or GNIS (U.S. Geological Survey 1981) but other gazetteers could be used as well. The database software programmed for the purpose was 4D (4th Dimension, San Jose, CA). It was customized to permit a combination of automated and manual matching of each specimen locality with place name(s) for the relevant county in the reference gazetteer.

Scripts that run from within the application find any matches in the reference database, using a simple system of pattern matching. Each entry in the GNIS for the county listed for a particular specimen is compared with the locality itself, and if any string of text in the locality description exactly matches a given entry in the GNIS, that entry is flagged as a potential match. Following this automated candidate match selection, the user can modify the selection by means of a simple interface. For example, when multiple matches for a particular locality occur, the user can select the most appropriate one with computer-mouse clicks. When no match is found, or if none of those found

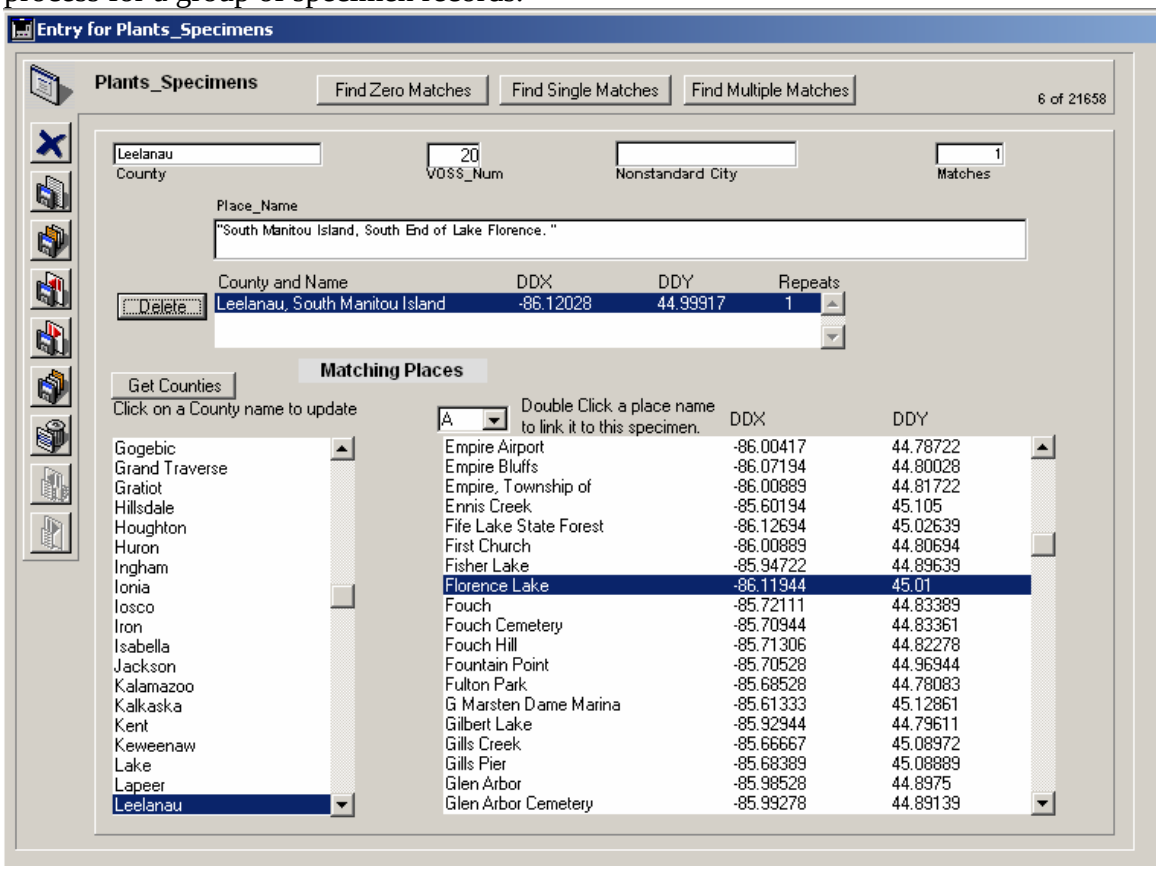
† This paper was posted November 2005 at DigitalCommons@University of Nebraska – Lincoln. http://digitalcommons.unl.edu/natrespapers/

automatically are appropriate, the user can easily retrieve the full list of places for the county, or other counties if desired, and then select an appropriate match. This feature is helpful for example when there are different versions or spellings of a place name.

To help illustrate the technique, a figure is included below showing the 4D display for one locality, 'South Manitou Island, South End of Lake Florence.' It has automatically been matched with the GNIS entry for 'South Manitou Island' in Leelanau County. The number '1' under 'Repeats' indicates that there was only one entry for that place name in Leelanau County. Had there been multiple entries, the user would need to determine which was the most appropriate one and select it. Lake Florence is included in the GNIS, but because the order of the words is reversed, no match was found automatically. The user has however recognized that coordinates for Lake Florence would be closer to the actual location at which the specimen was described as being selected. The user has therefore searched for and found the relevant entry, 'Florence Lake,' in the list of places for Leelanau County. Double-clicking on it will cause that entry to be added to the 'County and Name' area. By selecting South Manitou Island and using the 'Delete' button, that entry will be removed.

In a project in which this technique was used with herbarium specimens, a 92% success rate in georeferencing to some level of precision/specificity better than county was obtained (Simpson, ms).

Figure: Screenshot showing 4D display for one locality during the manual matching process for a group of specimen records.



Acknowledgements

Simpson was supported during this work by a joint postdoctoral fellowship from the University Library and the Rackham Graduate School, University of Michigan. Ed Voss of the University of Michigan Herbarium permitted use of specimen data from The Michigan Flora Project for the work.

Literature Cited

U.S. Geological Survey. May 1, 1981. Geographic Names Information System (GNIS). Reston, Virginia. Available online at URL: http://geonames.usgs.gov and http://nationalmap.usgs.gov.