Chabot: A System for Retrieval from a Relational Database of Images

Virginia E. Ogle
University of California, Berkeley
ginger@cs.berkeley.edu

Abstract

Chabot is a picture retrieval system for a database that will eventually include over 500,000 digitized multi-resolution images. We describe the design and construction of this system which uses the relational database management system POSTGRES for storing and managing the images and their associated textual data. For retrieval, Chabot uses tools provided by POSTGRES, such as representation of complex data types, a rich query language, and extensible types and functions. To implement retrieval from the current collection of 11,643 images, Chabot integrates the use of stored text and other data types with content-based analysis of the images to perform "concept queries".

1. Introduction

The Chabot project was initiated at UC Berkeley to study storage and retrieval from a large collection of digitized images. The images we use belong to the State of California Department of Water Resources (DWR), the agency that oversees the system of reservoirs, aqueducts and water pumping stations throughout California known as the State Water Project. DWR maintains a growing collection of over 500,000 photographs, negatives, and slides, primarily images of State Water Project facilities, but also many images of California natural resources. Some examples of these images are shown in Figure 1.

Over the years, as DWR has made its collection available to the public, it has found itself devoting increasing resources toward filling requests for prints and slides. The agency receives 100-150 requests a month from a variety of sources: other government agencies, regional magazines, encyclopedia, university libraries, wildlife organizations, and individuals. Requests vary from those where the ID number of the desired picture is already known, to very general requests for "scenic pictures" of lakes and waterways. DWR keeps the slides that are requested most often in lighted display boxes for browsing; the rest of the collection is