ECAI Infrastructure Activities at the University of California, Berkeley

Thomas Duncan
Museum Informatics Project
University of California, Berkeley
Museum Informatics Project

Goals

• Develop and implement an infrastructure to support deployment of electronic information systems in museums and archives
• Evaluate, select, and implement or develop data access, analysis, and information management software tools for scholars, curators, and collection management staff
• Operate a demonstration and development facility containing leading-edge hardware, software, and networking tools relevant to museum informatics
Operating Principles

• Data structures are more important than their interfaces
• Define an infrastructure based on open systems and standards
• Use commercially available tools to the greatest extent possible
• Infrastructure should be scalable, flexible, and modular
Purposes of Presentation

- Outline the institutional implications of projects such as ECAI
- Describe infrastructure work underway at Berkeley to support projects such as ECAI
- Describe the implementation strategies taken
- Outline results to date and plans for the immediate future
Institutional Issues vis à vis ECAI

- ECAI is composed of many individual projects that share a common infrastructure
- Multidisciplinary projects as ECAI require institutional support
- Institutions must identify and implement an information technology infrastructure to support projects such as ECAI
ECAI Infrastructure Components

- Client - server database support
- Repository services
- Software tools for spatial analysis
- Data entry / cataloging and retrieval tools
- Multilingual character sets
Client - server Database Architecture

- Client / server database architecture composed of three or more tiers; presentation, application and business logic, data
- Client applications used primarily for presentation to minimize application deployment and maintenance costs
- Applications composed of reusable components discoverable at runtime using Java and CORBA
- Component management at the institutional level
- At Berkeley:
  - data tier: Sybase
  - application and business logic: Jaguar CTM
  - presentation tier: Java applications or Web browser
Repository Services

- Creation of digital objects
- Archiving digital objects
- Accessing digital objects
- Repositories are for archiving and accessing digital objects
- Repositories are needed for many types of digital objects - digital images and structured text to be discussed
Repository Service Components

- Archival storage
- Registration authority
- Copyright management
- User authentication
- Access protocols
- Metadata management
- Data presentation
Repository Services: Images

• Image repository for museums, archives, and libraries under development
• A collaboration between Information Systems and Technology, The University Library, and campus museums and archives
• A testbed for the development of campus digital library services
• A prototype to support campus digital imaging projects
Image Repository Status

- Moving approx. 50,000 images to mass storage device
- Creating Sybase database of metadata specifications for images
- Creating a Sybase database of JTIP compressed images for presentation on the WWW
- Indexing and registering images through using software from Netimage as part of ISO JURA initiative
- Collaboration with Luna Imaging beginning for use of Insight as image browser
Prototype JTIP Viewer
Repository Services: Structured Text

- Support for SGML and XML documents as a searchable database
- Attempting to construct an integrated data management environment for structured data and structured text
- SGML structured text widely used in libraries and archives
- Related to multivalent documents to be described later in this meeting
Structured Text Repository Status

• Using full-text search engine developed by Verity and provided with Sybase
• Awaiting release of enhanced version of full-text search engine; may begin with pre-release copy
• Initial experiments with Korean Buddhist canon
• Search methods
  – Word, phrase matching
  – Proximity searches
Tools for Spatial Analysis

- Sybase supports spatial data types (points, polygons, lines)
- Geographic Information Science Center recently formed at Berkeley; discussion of needs relative to ECAI to be held following ECAI meeting
- ESRI site license to Berkeley completed
- Awaiting status report of Time Map project to determine how Berkeley will support the software and data structures recommended by that project
Cataloging and Retrieval Tools


- A thesaurus is a controlled vocabulary arranged in a known order in which equivalence, hierarchical, and associative relationships among terms are clearly displayed and identified by standardized relationship indicators.

- Used for consistent semantics when cataloging and to facilitate access

- Museum information systems utilize thesauri: e.g. taxonomic names and associated classifications, classifications of cultural terminology, and gazetteers of political, social, and biological geography
Representing Thesauri in an SQL Relational Database Management System

- Hierarchical relationships described by visitation model using
  - edge representation show broader / narrower term relationships
  - transitive closure representation is the distance from node (term) to root (top term)
  - visitation representation represents nested subsets of terms within thesaurus
- Associative and equivalence relationships specified through relationship types
- Multiple parents accommodated
Example Thesaurus

- Cultural terms
  - Art and Architecture Thesaurus
  - Library of Congress Subject Headings

- Biological classifications
  - Mammal Species of the World
  - Checklist of Vascular Plants of North America

- Political geography
  - Thesaurus of Geographic Names
Thesaurus Browser

Current Hierarchy:
- Earth
  - Western Hemisphere
    - North America
      - United States
        - California
          - Los Angeles
          - San Francisco

Matching Places:
- California

Related Concepts:
- Alternate Broader Term
- Pacific Rim
Multiple Parents for a Term
Multilingual Character Sets

- Unicode support now available in Sybase
- Implementing Academia Sinica database of missing CJK characters in Sybase
- Web-based interfaces to the database for missing CJK characters to be discussed following this meeting
- Planning for collaboration between UC Berkeley and Academia Sinica in this area