Academia-Sinica Digital Library: An Open Digital Archive Environment

D. T. Lee
Institute of Information Science
Academia Sinica
Taipei, Taiwan
Introduction

• American Memory (Library of Congress)’90
• G-7 Information Society and the Bibliotheca Universalis’95
• Memory of the World Programme (United Nations)’96
• Digital Libraries Initiatives I,II  NSF/DARPA/NASA.. ‘94-
  – make use of Information technologies to support creation, access and use of digital contents over the Internet
  – new communities of researchers, information providers and users have become engaged
  – vast amounts of digital information easily accessible to and usable by large segments of the world population
  – encompass not only science and engineering applications but also humanities and social sciences
  • Phase I: NSF/DARPA/NASA
  • Phase II: NSF/DARPA/NASA/NLM/LOC/NEH/FBI/NARA/SI/IMLS
Digital Libraries

• What is a digital library?
  – a library that maintains all or a substantial part, of its
collection in computer-processible form as an alternative,
supplement, or complement to the conventional printed and
microfilm materials (Bill Saffady 1995)
  – a collection of electronic resources, put together for access
and use by people over the Internet
  – a virtual environment comprising of digital collections in
which users browse them, use them to create new digital
information, and distribute them over the Internet
  – an environment to bring together collections, services,
and people in support of full life cycle of creation,
dissemination, use, and preservation of data, information,
and knowledge (1997 Santa Fe Planning Workshop)
Digital Libraries

• What is a digital library? System’s viewpoint

**It is a human-and-content-centered system**

- an extension of an information storage and information retrieval system that allows manipulation of digital data in any media such as text, image, audio and video
- allow sharing of a diverse range of collections of digital objects assembled by information content providers based on different principles
- allow collaboration and interaction among diverse people
- allow creation of information and knowledge.
Academia-Sinica Digital Library

Goals:
Build a large-scale, ubiquitously accessible, fully-integrated, human-centered digital library
• Support both learning and knowledge discovery
• Have added values to education, research and commerce
• Improve conduct of science, research and business
• Provide an excellent education facility to help create information-centered schools, universities or life-long learning institutions
ODAE Design Goals

Service Viewpoint

• Digital Archives Environment
  – User-centered, allowing for information gathering, storage, indexing, management, use and dissemination

• Metadata Management
  – Standardization (internationalization) to facilitate information exchange/sharing
  – Domain Knowledge of diverse people

• Multimedia Processing and Presentation

• Multilingual and missing character handling

• Information Assurance and authentication
## User Interface 2

### Presentation

<table>
<thead>
<tr>
<th>Hierarchical Browsing</th>
<th>Image Browsing</th>
<th>Video Streaming</th>
<th>QBIC</th>
<th>Full-text Searching</th>
<th>Web-page Searching</th>
<th>DB Searching</th>
</tr>
</thead>
</table>

### User Groups

- **General Users**
- **Advanced Users**
- **Content Providers & Info. Sys. Specialists**

**User Feedback, User-Profile Study & System Administration**
# User Interface

## Presentation

<table>
<thead>
<tr>
<th>Hierarchical Browsing</th>
<th>Image Browsing</th>
<th>Video Streaming</th>
<th>QBIC</th>
<th>Full-text Searching</th>
<th>Web-page Searching</th>
<th>DB Searching</th>
</tr>
</thead>
</table>

## General Users

- Metadata Management
- Database Management
- Multimedia Digital Collections
  (text, image, video & audio, etc.)

## Advanced Users

- Content Providers & Info. Sys. Specialists
- User Feedback, User-Profile Study & System Administration

December 4, 2001
PNC 2001 Mexico
ASDL System Components

Collections → Validation and Verification → Multimedia raw data and metadata

Collections → Back-end Production → Knowledge Discovery

Validation and Verification → Front-end Production

Front-end Production → Dissemination

Dissemination → Presentation

Front-end Production → Preservation

Preservation → Knowledge Discovery

Knowledge Discovery → Back-end Production

Other Archives Servers

Security and Rights Management | System Management | Session Management | User Service and Management | Broadband System Technology

December 4, 2001
PNC 2001 Mexico
System Design Issues

• **Scalability**
  – performance and effectiveness (e.g. CD vs Internet)

• **Interoperability**
  – syntactic and semantic Interoperability
  – system Interoperability
  – linguistic Interoperability

• **Extensibility**
  – Content creation and portability

• **Adaptability & durability**
  – Preservation, sustainability
Content Management Structure

- Content Manager
- Backend Media Storage
  - Multimedia File
  - FTP or Disc
- Search Engine
- ODAE Server
- Backend Media Server
- Export
- Static Pages
  - Video
  - Audio
  - Image
- External Web Server
  - Streaming Server
  - Search Engine
  - user#1
  - user#2
  - user#3
Digital Media Production

• Transform multimedia data into proper forms for efficient search and retrieval (Back-end production)
  – media compression
  – thumbnail generation
  – image stitching
  – scene/cut analyses
  – feature extraction
  – watermarking

• Present selective content from archives database (Front-end production)
Digital Library Kernel

• Effectively sort data into proper categories, and provide information management tools
• Define standard-compliant metadata
• Other management features
  – Rights management (IPR, access and copyrights)
  – Data backup and recovery
  – Software maintenance
Computer-Aided Tools

• Content distribution and multimedia streaming
  – VOD relay and streaming servers
• Easy-to-use computer-aided tools for content providers to present digital content
  – Augmented-Reality (-Panorama)
• Web-based database management facility
ASDL Database Management

Back-end Production---Support the following:

• Archives Database Creation/Management
  – Metadata Definition, Database Schema
  • Web-based user interface for data acquisition
  • DBMS support

• XML based Data Exchange
  – Simple Dublin Core, Dublin Core, etc.

• Multimode Searching
  – Metadata, Dublin Core and Full Text

• Multimedia Presentation
  – Thumbnail, Browsing
  – (Video, Audio, Image)-on-Demand
## Metadata for video tapes of the Pepo Project

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata 項目分析與分類表</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster</td>
<td>Core Class</td>
<td>Object</td>
<td>影像</td>
<td>系統原則</td>
<td>參考原則</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>入門</td>
<td>影像在內容上所指涉時間</td>
<td>*設定理由（可選選）（1）事前</td>
<td>*區隔使用編號</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2</td>
<td>時間</td>
<td>影像在內容上所指涉時間</td>
<td>*設定理由（可選選）（1）事前</td>
<td>*區隔使用編號</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>人物</td>
<td>影像在內容上所指涉重要地點，聚落</td>
<td><em>姓名表示方法：今（古），例：</em></td>
<td>*區隔使用編號</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>關係</td>
<td>影像所指涉之單位描述名稱</td>
<td>*區隔使用編號</td>
<td>*數量不限</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>5</td>
<td>事件</td>
<td>影像所指涉之單位描述名稱</td>
<td>*區隔使用編號</td>
<td>*數量不限</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>6</td>
<td>討論</td>
<td>影像內容分類還之內容</td>
<td>*區隔使用編號</td>
<td>*數量不限</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>7</td>
<td>開頭</td>
<td>影像內容所指涉討論</td>
<td>*區隔使用編號</td>
<td>*數量不限</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>8</td>
<td>物</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>9</td>
<td>摘要</td>
<td>研究人員針對影像內</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>10</td>
<td>著作權</td>
<td>影像的保存</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>11</td>
<td>採訪地點</td>
<td>影像的保存</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Entry spec.**
Database Attribute Management

- Field Name
- DC mapping
- Attribute Type
- Indexing

Archives Type
Video Tapes
Historical Map
Add New Video Title

Heading

Subject

Keywords

Abstract

Video Tapes
Search Interface 3
Media Server/Client Requirement

• Media Server:
  – CPU: Pentium II-300 & above
  – RAM: at least 256MB RAM
  – DB: MS SQL Server (1 License)

• Media Client:
  – CPU: Pentium 200 & above
  – RAM: at least 128MB
Multimedia on Demand System

Front-end production

- Multimedia on demand, including video, audio, images, CD titles, etc.
- Media Searching
- Direct Browsing, Retrieval and Playback
- Supports AVI, MPEG-1, DAT, Quicktime, MP3, WAV, etc.
Distributed MM System

• Management/Maintenance done via Web Browser
• User authentication
• Real-time bandwidth monitoring and control
• Individual or group access control per media license agreement
• Accounting & bookkeeping of user and media usage
Distributed Media Servers

- Media Server Group: Media servers can be interconnected to provide resource sharing
- Transmission Relay Mechanism: high throughput QoS
Distributed Media Databases

• Web-based interface for management of all distributed media databases
• Manager of each server cluster aware of the status of each media database
• Download media to selected media server cluster
<table>
<thead>
<tr>
<th>title</th>
<th>format</th>
<th>language</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistics
media
### Past record

<table>
<thead>
<tr>
<th>姓名</th>
<th>登入時間</th>
<th>離開時間</th>
<th>IP位址</th>
</tr>
</thead>
<tbody>
<tr>
<td>管理者</td>
<td>2001/7/12 下午 12:11:30</td>
<td>2001/7/12 下午 12:22:17</td>
<td>140.109.18.168</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/12 上午 10:27:54</td>
<td>2001/7/12 上午 10:33:13</td>
<td>140.109.18.123</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:33:48</td>
<td>2001/7/11 下午 09:35:03</td>
<td>140.109.18.99</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:28:49</td>
<td>2001/7/11 下午 09:34:48</td>
<td>140.109.18.100</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:33:52</td>
<td>2001/7/11 下午 09:25:42</td>
<td>140.109.18.165</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:22:20</td>
<td>2001/7/11 下午 09:23:42</td>
<td>163.27.31.10</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:08:19</td>
<td>2001/7/11 下午 09:23:52</td>
<td>140.109.18.99</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:05:42</td>
<td>2001/7/11 下午 09:08:19</td>
<td>140.109.18.99</td>
</tr>
<tr>
<td>管理者</td>
<td>2001/7/11 下午 09:02:22</td>
<td>2001/7/11 下午 09:03:56</td>
<td>140.109.18.99</td>
</tr>
</tbody>
</table>
Image Server: RiverImage

• Current Status
  – ODAE supported projects in Academia Sinica
    http://odae.iis.sinica.edu.tw
    http://www.sinica.edu.tw/~pingpu
    http://mip.iis.sinica.edu.tw/arthistory
  – National Palace Museum (http://npmpc.iis.sinica.edu.tw)
  – National University of the Arts
    (http://dialog.iis.sinica.edu.tw)

• Features Customizable presentation interface
  – Support Multiple Image Formats
  – On-line browsing of large photos/images
Watermarking: Image Protection

- Visible watermark is plainly visible across the body of the image or situated on the side.
Invisible Watermarking—Robust Cocktail

- Blurring
- JPEG 5%
- Histogram equalization
- Sharpening 85%

December 4, 2001
PNC 2001 Mexico
Invisible Watermarking - Fragile

original image  watermarked image  altered image

2nd level  3rd level  4th level
• List of DigiLock (www.digibitsinteractive.com) Customers and Evaluators
  – National Palace Museum
  – Asian Art Museum of San Francisco
  – Luna Image
  – Lee&Lee Communications
  – IMAT (Int’l Multimedia Arts & Technology)
  – Comstock
  – PPA (Prof. Photography Assoc.)
  – Bluce Coleman, etc
Take a Closer Look at the Original (L) and Watermarked (R) Frames
Video Watermark Detection Result with threshold determined Under False Positive Prob. $10^{-6}$

Detection results of compressed video watermarks

- Flower-Garden (watermarked)
- Flower-Garden (sharpening)
- Flower-Garden (frame-rate changing)
- Flower-Garden (MPEG: 6 Mbit/s)
- Flower-Garden (MPEG: 4 Mbit/s)
- Flower-Garden (MPEG: 2 Mbit/s)
- Flower-Garden (noise adding)
- Flower-Garden (frame averaging)
- Flower-Garden (mean filtering (3x3))
- Susi (non-watermarked)
- Mobile (non-watermarked)
- Table-Tennis (non-watermarked)

Detection value vs. I Frames

Threshold
Search Engine Technology

• Input Data Types
  – Database Tables
  – HTML and BBS Documents

• Search Functions
  – Full-text Search
    • for general user
    • better recall rate
  – Metadata Search
    • higher precision
    • 15 elements of Dublin core
    • simplified version
Conclusions

• Built an open environment for content creation, use and management
• Established a multimedia data processing facility
• Defined workflow of media production process
• Developed core technologies
  – Web-based metadata & database management interface
  – Distributed media server & media client
  – Watermarking for media protection
  – Search engine for Chinese
• http://odae.iis.sinica.edu.tw/
Thank you for your attention!

謝謝大家！

Muchas Gracias!