投影片摘要

- Content Management of Academia-Sinica Digital Library
Content Management of Academia-Sinica Digital Library

Jan-Ming Ho and D. T. Lee
Institute of Information Science
Academia Sinica
Taipei, Taiwan
Contact author: dtlee@iis.sinica.edu.tw
Digital Libraries

  - 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

- Digital Libraries Initiatives Phase II (1999 - )
  - Encompass not only science and engineering applications but also humanities and social sciences
  - NSF/DARPA/NASA/NLM/LOC/NEH/FBI/NARA/SI/IMLS
Digital Libraries

- What is a digital library?
  - 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)
  - 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 collaboration and interaction among diverse people
    - allow creation of information and knowledge within it.
Academia-Sinica Digital Library (I)

- Computer Technology and Industry of Taiwan-ranked 3rd (next to U.S. and Japan), & NII in place in 1994
- Decent contents lacking
  - e.g., educational content on science, history, culture
- Ubiquity of Internet
- Content Industry
- Project officially initiated in Dec. 1998
- National Digital Library/Archive Project in 2001
Acadmia-Sinica Digital Library (II)

- 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
- Provide an excellent education environment to help create information-centered schools, universities and life-long learning institutions
User Interface

Presentation

General Users

Advanced Users

Metadata Management

Database Management

Multimedia Digital Collections or Archives

Hierarchical Browsing

Image Browsing

Video Streaming

QBIC

Full-text Searching

Web-page Searching

DB Searching

Content Providers & Info. Sys. Specialists

System Administration, User Feedback & User-Profile Study

January 15, 2000
Content-based Image Retrieval

Query Results

Image query

Text query

Hyperlink to other database

January 15, 2000
Content-based Image Retrieval

Query Input

- Static Image
- Dynamic Image
- Uploaded Image

Search Engine

- Feature Vector Database
- Image Database
- Related Link Database

Search Related Information

Query Results

January 15, 2000
System Design Issues

- Scalability
  - performance and effectiveness (e.g. CD vs Internet)
- Interoperability
  - multiple data sources, heterogeneous objects
  - system Interoperability
  - linguistic Interoperability
- Extensibility
  - Content creation and portability
- Adaptability & durability
  - Preservation, sustainability

January 15, 2000
System Architecture

Collection → Backend Production → Preservation → Front-end Production → Dissemination → Presentation

Validation and Verification

Knowledge Discovery

Other Archives Servers

Multimedia raw data and metadata

<table>
<thead>
<tr>
<th>Security and Rights Management</th>
<th>System Management</th>
<th>Session Management</th>
<th>User Service and Management</th>
<th>Broadband System Technology</th>
</tr>
</thead>
</table>

January 15, 2000
Digital Media Production

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

• Publish 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
- Fast database generator
VOD Server

Terabyte HSM

Encode Server

TCP Streaming Server

UDP Streaming Server

ADSL Streaming Server

Intranet

Broadband Internet

ADSL

January 15, 2000
Database Generator (I)

- Enable user to create a database through a Data Description Language (DDL)
- Support database management and queries on WWW
- Core system consists of toolbox on browser, Intermediate server, DBMS
Database Generator (II)

Example
Metadata – Types

- Person
- Object
- Time
- Space

Events connect these entities in a circular diagram.
Metadata - Record Structure

CORE

INDIVIDUAL

CDWA, CIMI, DC, EAD, FGDC, NGDF, TEI etc.
Watermarking - Image Protection

blurring

JPEG 5%

histogram equalization

sharpening 85%

central研究院

January 15, 2000
Watermarking - Image Authentication

- Original image
- Watermarked image
- Altered image

2nd level
3rd level
4th level

January 15, 2000
Server Architecture

Content Management Structure

January 15, 2000
Digital Audio Production Workflow

1. Backend Production Client
2. Meta data server
3. Audio Repository Manager
4. Exporter

WAV audio

Audio

Server ID
File ID

Audio meta

Server ID
File ID

Audio meta

January 15, 2000
Enter object identification

Content production

Article

Media

Map
請輸入物件識別碼

識別碼 type: 個人編碼
識別碼: test001
單張地圖  Single Sheet Map

Input scanned map
3 columns by 2 rows, file format: TIFF

1. & 2. Enter the dimension of the map
3. Enter its scanned file format
SetNumber there are 3X2 images
mappath=../../htdocs/dlimages/map/

1. Enter the image file names in each field
2. Assign output file name—Title (a3x2)
Back to the original input interface
Output file name is recorded
MAP

物件識別類型： URL 識別碼： abc001
題名： a3x2
檔案名稱： a3x2

- 開放給所有人查詢
- 僅開放給本群組查詢
- 暫時不開放查詢

Sent to Archives Database  Back for further revision
若要產生HTML (static page) 以供搜尋，請按下面的 BUTTON 《產生》
### Query

[目前是第1頁/共72頁/找到573筆資料]

<table>
<thead>
<tr>
<th>笔数</th>
<th>識別碼type</th>
<th>識別碼</th>
<th>TITLE</th>
<th>繪製者</th>
<th>處理狀態</th>
<th>細目</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>map</td>
<td>test006</td>
<td>ss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>URL</td>
<td>abc001</td>
<td>a3x2</td>
<td></td>
<td>處理完畢</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>URL</td>
<td>adad</td>
<td>adad</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Status check:** If processing is complete, one can check the details by pressing “Itemize” button

January 15, 2000
This gives a summary of the metadata previously entered

Clicking at the thumbnail will get a bigger picture

January 15, 2000
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 environment for content creation, use and management
  - Defined workflow of media production process
  - Developed core technologies
    - Fast database generator
    - Audio and video streaming servers
    - Watermark
    - Domain-specific QBIC
    - Search engine for Chinese
  - http://dl.iis.sinica.edu.tw/
- Characteristics of Digital Library
  - Basic research and engineering development
  - Multi-disciplinary
    - collaboration and interaction among diverse people
Future Work

- Data Model and System Analysis, Design and Implementation
- Field Data Collection Database Design and Implementation
- Distributed Archives Database Design and Implementation
- Integration of Multimedia Contents, Metadata, Knowledge on Multimedia Processing Tools, and System Management Tools
  - the above four items will be conducted in cooperation with the other institutes participating in the National Digital Archive Project
- Computer Aided Categorization of Multimedia Objects
- Performance Optimization of Integrated Browser and Search Engine for Multimedia Objects
  - e.g., scalability problem, re-index problem, processing of compound queries
- Distribution and Streaming of Multimedia Contents
  - e.g., QoS Guarantees and system scalability