Cultural Heritage Retrieval

Jaap Kamps
University of Amsterdam
http://staff.science.uva.nl/~kamps/

Multimediale Information Retrieval
Utrecht, January 12, 2009
Outline

• Motivation
  ⭐ CH Websites are incunabula

• State of the Art
  ⭐ What’s in a museum?
  ⭐ How to get it out?

• Required changes
  ⭐ In four slogans
Motivation: CH Websites are Incunabula

(Biblia Latina, 1477, UC Riverside)
Incunabula

- **Incunabula** are the earliest printed books that mimic the hand-written editions
  - complete with coloring and illuminations by hand
  - the printing press is used to improve the earlier manual book production process
  - rather than make full use of its potential to mass-produce books

- Websites of CH institutions are currently “incunabula”
  - they do not take full advantage of the (technical) possibilities
  - fail to maximize their mission as memory-institutions of the Netherlands/Europe/World
State of the Art: Museum Documentation

- The **Gemeentemuseum, The Hague** combines all three CH traditions:
  - **Museum**: 128,000 museum object descriptions
  - **Library**: 279,000 bibliographic descriptions (books, pamphlets, articles, etc.)
  - **Archive**: 729,000 process related descriptions

- Data in three different, tailor-made systems
  - You cannot search across all information
  - Much data is administrative, little on content
Traditional Systems – User (Un)friendlyness

- Many tabs, hard to find information, easy to get lost, most fields are never used.
Some Recent Developments (1)

- Getting the CH data is always hard...
  - Getting permission in the first place
  - Usually locked inside proprietary systems
  - Format differs, semantic interoperability

- Perhaps “Freedom of Information” act may help?
  - Won’t allow re-publication though...

See: Frankie Roberto, Exploring Museum Collections On-line
Some Recent Developments (2)

- Presenting CH data is hard...
  - Especially if its in large volumes
  - from different sources

- Dual problem
  - Data/semantic interoperability: data formats are incompatible
  - User interoperability: out-of-context data is uninterpretable

- Much effort in “controlled vocabularies” both popular in traditional CH and Semantic Web

See: Johan Møhlenfeldt Jensen, Approaches To Presentation Of Cultural Heritage Information
Some Recent Developments (3)

- Searching across collections is hard...
  - Several experiments with federated search
  - i.e., each institution offers content through a web service

- However, effective search is key
  - and usually an afterthought!
  - For effectiveness a central index is far superior
  - For efficiency a central (mirror) index is unavoidable

See: Terry Makewell, The National Museums Online Learning Project Federated Collections Search
Required Changes

- What is needed to unleash the power of the web in CH?
  - Still largely an open question
  - We need help!

- Four slogans
  - *Open the depot*
  - *Don’t patronize*
  - *Art is all about interpretation*
  - *Take users seriously*
(1) Open the Depot!

(Schaudepot, Museum Rietberg, Zurich)
Provide Access to All Objects...

- Only a fraction of museum objects are on physical display
  - often not much more on their websites

- Yet usually descriptions and images exist in digital form
  - gM website: 1,130 objects
  - gM databases have hidden gems: 128,000 object descriptions; 279,000 bibliographic descriptions; and 729,000 archival item descriptions.

- How to make all (digital) descriptions/images available in any shape or form, effectively part of the virtual collection?
Demo System: CatchUp

- MuSeUM project build a search engine for all data
  - Reusable evaluation sets to determine quality
  - Improving search technology for museum data, and requests at hand (differentiating between expert and naive users).
  - [http://staff.science.uva.nl/~kamps/museum/demo/](http://staff.science.uva.nl/~kamps/museum/demo/)
(2) Don’t Patronize!
Don’t Patronize...

- Traditionally curators decide what the public get to see, and how to interpret it
  - Modern users want to decide for themselves,
  - across the boundaries of institutes and collections, and across traditional views on art?

- Need to make a fundamental switch
  - Glass is half empty: don’t publish anything unless “perfect”
  - Glass is half full: publish anything with potential value
  - Of course, with the right contextual information...

- This is the main barrier: it requires a complete change of philosophy of curation and curators!
(3) Art is All about Interpretation!
Art is All about Interpretation...

- Explore what’s available from many angles
  - Rather than the straightforward one-step communication of factual information,
  - searchers want to explore the collection by searching and browsing and create their own story.

- This requires highly interactive search
  - support rich means of search and navigation,
  - specific to the search trajectory (and user profile)
  - from any object access similar objects by creator, style, time-period, recommendation, etc.
Demo System: MultiMatch

- EU project on digital Cultural Heritage
  [http://multimatch.org/](http://multimatch.org/)

  - Aims to enable users to explore and interact with online accessible cultural heritage content, across media types and languages boundaries.
  - Combining traditional information (libraries, archives, and museums) with information on the Web at large (Wikipedia, newsfeeds, blogs, etc).
  - Main idea: instead of providing a complex summary page of all information in all sources/languages/media-types, you let a searcher navigate this “space”
(4) Take the User Seriously!

(gemeentemuseum.nl) (nl.wikipedia.org)
Take the User Seriously

• Traditionally, CH was in libraries, archives, and museums
  ★ This is now complemented by a massive amount of information in consumer-generated resources as Wikipedia or blogs.

• How to combine this multitude of perspectives on CH content?
  ★ MuSeUM project is setting up a collaborative site
  ★ users can rate/bookmark/tag objects, leave comments and discussion, create and share virtual exhibitions, play games (ordering objects by date, price, same painter), etc.
  ★ Why not http://del.icio.us/? we want to preserve the local community of the museum
Demo: My Museum

- [http://my.gemeentemuseum.nl](http://my.gemeentemuseum.nl)

  ★ Actually broader aim: it is an experimental environment for “new media” approaches
  ★ Allows for experiments, on real data, with new approaches
  ★ Successful experiments can be promoted to the real website
  ★ First experiment *gemeén*: mostly internal discussion on “new media” in the museum
  ★ Second experiment *MyMuseum*: collaborative site for school children visiting the museum

- What more can we do? We’re open to suggestions!
Wrap up

- CH web sites are incunabula
- State of the Art
- Four slogans
  - *Open the depot*
  - *Don’t patronize*
  - *Art is all about interpretation*
  - *Take users seriously*
- Request for ideas on using the experimental platform
Thank you!

Questions?