Constructing a digital museum with a large-scale archive for endangered languages

Yukinori Takubo, Tamaki Motoki, Syuntaroo Tida, Shoji Kajita, Masahiro Yamada, Yoshihiko Asao, & Keisuke Yagi

info@kikigengo.jp
Goals of the Museum

• To design a prototype of the web-based museum for endangered languages and cultures that is easily updatable, and serves as a basis of collaborative research.

• To provide a forum in which the local people can exhibit their language products.

• To encourage the local people to have a wider access to language resources and help them to preserve the language and culture of the local community.
Design of the Museum

• Four layered web-based museum
  – Open space:
    • The exhibit space
    • Library and references
  – Closed space:
    • Closed library and archives
    • Data storage space
Spaces with closed access

• Closed library and archives:
  – Accessed only by closed members.
  – Files with transcription and other updated archiving information are stored here.

• Data Storage Space:
  – Accessed only by our research group.
  – Raw data files with basic meta-data are stored here.
Why ‘museum’?

- Metadata updating through exhibition at the Museum

**Open Access**
- The exhibit space
  - Contents with various interfaces
- Library and references
  - Contents with updated metadata

**Closed Access**
- Closed library and archives
  - Data, Metadata, Information
- Data storage space
  - Files with Metadata

---

**Making contents for exhibition:**
Transcription, Translation, movie edition, adding comments

---

**Exhibition of the contents**

**Store used contents with metadata added**
Our previous work

• Prototype of our Digital Museum: Ikema-Nishihara
  – Metadata updating through exhibition
  – Movie contents with subtitles
  But
  – No online database. External hard disk drives were used to store data.
  – Sharing of data was done offline.

Cf. Takubo, Yukinori (forthcoming), Takubo, Y and T. Motoki
http://Kikigengo.jp
Our previous work

- The database is not designed for constructing data archive. It is, therefore, difficult to manage individual files.
New feature: Adding a large scale online database

Open Access Exhibit Spaces
- Exhibit space
  - Contents with various links, stories
- Library and references
  - Contents with updated metadata

Online Data Archives
- Data storage spaces
  - Data files with metadata
  - Various levels of access control

Creating contents for exhibition:
- Transcription
- Translation
- Movie editions
- Adding linguistic or cultural comments

Store contents with metadata added
Implementation before

Digital Museum
Exhibit Space
Library and Reference

Webpage and database intermeshed

Transfer
Transfer
Transfer
Transfer
Transfer
Transfer
Kaltura: An open source online media management system

• Publishers
  – can create channels for various purposes, e.g. languages, research groups, etc.

• Media management platform
  – account management, e.g. role restriction
  – Fine-grained access control
  – Media management, media players, streaming
  – Allows various search: descriptions, tags, categories, references
Implementation after

Digital Museum

A
Exhibit Space

Library and Reference

Link

Link

Link

Link

Environment

Web Browsing

Media Management System

Publisher α

URL

XML

Publisher β

URL

XML

Transfer

Transfer

Link

Link

Link

Link

Link

Link

Link

Link
Kaltura: Access control by Publishers

- A registered member of a publisher A can have access to the media files and the metadata in A, depending on the assigned role of the membership.
  e.g. can upload, edit, or only view and get a link URL of a media file
- Non-members cannot access the contents of a publisher.
Kaltura: Entries
Access control

- Closed access by signing into a Kaltura publisher
  - Access control for/over
    - membership (admin, editor, viewer etc.)
    - individual media files or categories over files
Kaltura: File Search

- IDs
- Categories
- Key words
- Tags
Current system of the Museum

Open Access Exhibit Spaces

- Dokuwiki
  - open source wiki
  - well maintained
  - various plug-ins provided
  - simple language and friendly interface

Online Data Archives

- Kaltura
  - media databases
  - media players
  - streaming

Media files stored at the data archive are embedded (via web links) at a dokuwiki page
Comparison with YouTube

• Can maintain the copy rights of the contents for
  :privacy protection,
  :limiting access to sensitive contents such religious rituals.
  :unwanted editing
• Fine-grained and fine-tuned access control.
Some Limitations

• No workshops: editing of media files must be done offline
• Search is by exact match only
• Still at an experimental stage
• Cannot replace real database (at least for the time being)
Some applications

• Yonaguni Project: linguists train local people to record and transcribe their speech, or speech by older people

• Miyako High School Project: High school students interview their grant parents and upload the files to a publisher.

• Miyako Project: native speakers of various Miyako dialects make recordings of their own speech and upload the files themselves.
Summary

Large scale online archive:

• Easy to install and set up
• Separate from the exhibition website
• Fine-grained access control
• Useful for the documentation and preservation of endangered languages
References

• Kaltura
  http://corp.kaltura.com/
• Kikigengo.jp
  https://www.facebook.com/kikigengojp

  Link downloadable from academia.edu

Demonstrations

- Kikigengo.jp
  - wordpress
  - dokuwiki
- Database page