An OAI-PMH Based Thumbnail Capture & Delivery Service

Timothy W. Cole, Thomas G. Habing
University of Illinois at Urbana-Champaign (USA)

Muriel Foulonneau
Centre National de la Recherche Scientifique (France)

Motivation:

Many OAI data providers do not maintain thumbnails. Others have thumbnails but do not include URLs for thumbnails in records disseminated via OAI-PMH.

Research suggests thumbnails are useful and facilitate identification & selection of resources by end-users.

UIUC thumbnail capture & delivery service enhances usefulness of our OAI-based portals. Also a model of more advanced OAI-mediated services, for example, enabling more repositories to participate in the DLF-Aquifer Asset Action experiments.

Capturing thumbnails:

THUMBGRABBER APPLICATION:

Retrieves resources over the Web, then either:

• Selects largest image on Web page, OR

• Captures Web browser window (800x800 pixels).

Image resized and saved as PNG file. Metadata added to the PNG file, including OAI identifier & source URL.

Application configurable using interactive GUI, at startup using command line options, or through reference to ancillary configuration files.

RELIANCE ON OAI-PMH:

ThumbGrabber reads harvested OAI records to find URLs of resources for which thumbnails are needed. Operates in BATCH mode, processing all OAI records in a given directory, or in ON-DEMAND mode, processing requests for thumbnails via a Web Service, and accessing OAI records directly from repositories using OAI GetRecord.

Delivering thumbnails:

3 SERVICES (may reside on separate servers):

Service 1: Maintains an index (currently SQL) tying OAI identifiers to the URL of corresponding thumbnail images. URLs can be distributed across many hosts.

Service 2: Stores thumbnails created by ThumbGrabber and makes them available via HTTP GET at persistent URLs. If no thumbnail available, returns single transparent pixel and 404 (not found) HTTP header.

Service 3: Processes ON-DEMAND requests: An HTTP POST contains OAI identifiers which are resolved to OAI Records using UIUC OAI Registry and added to a queue. Images are captured as described above.

Looking ahead:

DLF-Aquifer (Asset Actions) and OAI-ORE based projects will require specific views of resources. Our approach for thumbnails may provide a way forward for repositories not able to independently supply all required views; currently experimenting with prototypes of such services.

thumbnail requested by Web form (on demand)

Post ThumbGrabber Request

(authentication required)

OAI-PMH GetRecord

The baseURL for oai-identifiers that follow OAI implementation guidelines are resolved using the UIUC OAI Registry. Alternatively, the baseURL may be included with the ThumbGrabber request.

OAI identifier resolution service

bulk OAI harvesting & batch processing

OAI-PMH Records on disk

OAI-PMH harvester

XSLT processing

URLs from OAI metadata records

Metadata Aggregation

SQL, DLXS, ...

ThumbGrabber application

Open URL in Web Browser

Embedded Instance of MS IE

Browser Cache

capture Web page view — OR — capture largest image

screen capture window

retrieve image from cache

convert image to PNG format

add image metadata to PNG file

configuration and control parameters set through GUI, cmd line, and ancillary files:
• collections.xml to set by collection
• blacklist.xml excludes capture of logo & branding images used by some sites.

configuration & control

metadata added to PNG:
• title
• description
• OAI identifier
• URL source
• original width
• original height
• original MIME type
• original last modified
• software version

index of thumbnails by OAI identifier (incl. external images) *SQL*

thumbnail PNG images created by ThumbGrabber *HTTP*

Search & Discovery Service

ILLINOIS HARVEST

external thumbnail images from data providers

Thumbnail delivery service

search results page with thumbnails

Related projects:
• Illinois Harvest: http://illinoisharvest.grainger.uiuc.edu/
• IMLS Digital Collections & Content: http://imlsdcc.grainger.uiuc.edu/
• CIC Metadata Sharing Collaboration: http://cicharvest.grainger.uiuc.edu
• DLF-Aquifer Asset Actions: http://rama.grainger.uiuc.edu/assets

See also:
• Foulonneau, Muriel, Thomas G. Habing, and Timothy W. Cole. 2006. Automated Capture of Thumbnails and Thumbshots for Use by Metadata Aggregation Services. *D-Lib Magazine* 12 (1): doi:10.1045/january2006-foulonneau.
• Chavez, Robert, Timothy W. Cole, Jon Dunn, Muriel Foulonneau, Thomas G. Habing, William Parod, Thorton Staples. 2006. DLF-Aquifer Asset Actions Experiment: Demonstrating Value of Actionable URLs. *D-Lib Magazine* 12 (10): doi:10.1045/october2006-cole.
• API description: http://gita.grainger.uiuc.edu/registry/thumbnails/TGWS_API.htm

template to retrieve thumbnails:
http://servicebaseURL?identifier=oai-identifier

example:
http://gita.grainger.uiuc.edu/registry/thumbnails/thumbnail.asp?identifier=oai:lib.uchicago.edu:AEP-ILP69

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN