

Software Architecture Meeting Minutes

March 3, 2011

Present: Kalaivani Ananthan (recorder) Isaiah Beard, Jie Geng, Dave Hoover, Rhonda Marker, Chad Mills (Co-chair), Sho Nakagama, Jeffery Triggs, Yang Yu

Agenda:

- 1) Update
 - Google Scholar (Triggs)
 - R5.2 development status (All)
 - Shibboleth in production environment (Hoover)
 - PDF server/OCR of PDF issue (Mills)
- 2) Object updates - OCR datastreams and thumbnails (Triggs, all)
- 3) Planning JPE migration (Ananthan, all)
- 4) Handle server specification (Nakagama)
- 5) Handling modification requests (Ananthan, Mills)

1. Updates

Google Scholar

Jeffery reported that there were 1490 objects in Google Scholar on March 2, 2011 but then lost 60 of them this morning. He does not know why this might have happened. It could be because of some objects that we requested they take down. They are coming very frequently to get the site maps and may be close to the complete refresh. Jeffery will write up a script to analyze what objects are in Scholar at a given time, and Rhonda will discuss with Jeffery offline to figure out how to analyze the objects.

Browse by Publication Date list has been added to RUCore Search Portal on the production server and Jeffery reported that the list has already been picked up by Google Scholar.

5.2 development status

Yang reported that the development of WMS is on a tight schedule. He informed that the collection metadata elements have been added but he needs to verify that all the elements have been added. He is currently fine tuning the Cataloging Utility to ensure that everything is working OK. The WMS bugs still need to be fixed.

Jie is OK with the schedule. The Faculty Deposit Module is ready to test on the development site. XACML class is also ready for testing in dlr and WMS. Jie is working with Chad to trim the stats table. Stats will be migrated to a separate database in R 5.2.

Jeffery reported that the new showfed.php is ready to test as nparsefed.php. The showfed display will look similar to RUcore full record display. He will add a "Help" link to the new dlr man pages protected by dlr authentication level 1. Jeffery will start working on the MARC enhancement requests. Kalaivani will work with him to prioritize the requests, if needed.

Chad reported that he is on a tight schedule, though the majority of R5.2 development is done. He will be focusing on completing development of Annotation Tool for the next two weeks. The Annotation Tool can still be delivered to NJVid without FFMPEG compiled since Sujay already has it running. Sho will work with Dave to see if FFMPEG can be compiled with minimal configuration. Chad also reported that he is developing a new search API to extract datastreams for R5.2. This may cause the log files to grow too large. We can't change stats till Ron returns. Chad will send a URL to the development portal code for Dave and Sho to look at the entries in the log files.

Shibboleth

Dave reported that Chuck did his part and asked Dave for metadata. It is generating a certificate mismatch error. Dave is trying to isolate this problem. Sho thinks it could be the fact that the certificate was not cloned from Chuck's machine as it was with lefty64. Dave may ask for such a copy. The group discussed whether we want to use a Rutgers-only server for JPE video. Dave said that using a Rutgers-only server would create problems with users trying to access NJEdge videos.

Dave brought up the issue with space needed for indexing certain large collections. The Graduate School of NB is possibly the largest collection now, though the largest file is still only 900 MBs. Rhonda and Isaiah reported that they have deleted some NJEDL objects from WMS and from temp upload area.

PDF server/OCR of PDF issue

Production server is now being pointed at the test PDF server, which was working. Chad added two new attributes (mode=all and PDF conversion=monochrome) and the production machine started working again. The servers are currently reversed, with the functioning production server at the SCC and the TSB server acting as the test machine. We will continue to test in this environment and eventually switch back in a next release.

2) Object updates - OCR datastreams and thumbnails

Jeffery ran reports on all three servers and emailed the XML reports. He will separate the DjVu files and then identify PDFs that might usefully have OCR texts added. Jeffery will work with Linda Langsched to perform some test searches on the test server where the files have already been added. The group discussed about extracting text layer for handwritten manuscript. Jeffery will select a handful of files with garbage text to see how useful it would be to index them.

Thumbnails will be considered later as an enhancement. There is an outstanding bug related to thumbnails generation in WMS. Once this bug is fixed, the WMS will generate appropriate thumbnails and this will no longer be an issue.

3) Planning JPE migration

Kalaivani produced generic document outlining the migration process using an earlier version drafted by Jeffery for NJVid video objects migration. . Kalaivani will update the document with specific reference to the JPE issue. We need to consider how the Darwin server will be affected by the migration of video objects. We might be able to begin the JPE migration, but we will wait to hear from Dave before proceeding with the actual migration.

Rhonda would like to bring up the issue of large files. These are becoming more and more common. Fedora 3.5 will supposedly have more robust large file handling. This will be added as an agenda item for the next meeting.

Next meeting will be on 17th at 10 am. Isaiah will be the next minute recorder.