

Software Architecture Meeting, Thursday, December 13th, 2007

(Ananthan, Beard, Ellis, Geng, Hoover, Jantz, Liew, Marker, Mills, Nakagama, Triggs)

Agenda

- 1) PDF Server class, SOAP; Ellis, Mills.
- 2) Collection statistics update, collection stats audit and load issue on mss3; Geng, Mills, Triggs.
- 3) Continue Google indexing discussion - we need PID on the 'showfed' display
- 4) Checking of handles on MSS3 report, audit of handles; Triggs.
- 5) XACML follow up; Ananthan and Triggs.
- 6) Discussion with Fedora architect and impact on our release plan
- 7) Periodic audit of signatures on mss3 - are we doing this and who gets alerts?

General Announcements

Jantz will distribute a recently written Northwestern document to the group that makes substantial reference to RUCore infrastructure methodologies.

Next Meeting is Thursday, January 10th, 2008 in the Heyer Room at the SCC in Alexander Library.

Jantz sabbatical ends on January 1, 2008 and will be returning fulltime.

At the end of April '08 Fedora will release version 3.0, it will not include large file support. Large file support will be added in release 3.1, coming three months after the 3.0 release.

Agenda Items

- 1) PDF Server class, SOAP; Ellis, Mills.

Mills reported that the test PDF server is functioning as expected and a majority of the PHP class has been written and tested that will allow for the remote submission of "Office Documents" to be converted. The class uses SOAP to make the remote requests to the PDF server. In particular nusoap is being used, however with the addition of PHP5 in the future the class will be overhauled to use PHP5's built in SOAP functionality. Development of the class was done on "lefty." It was decided that metadata embedded in the source office document (Author, Subject, Title and Keyword) will be expunged during the creation of the PDF and will not appear in the final product. Security settings for generated PDF will remain at a minimum and not allow complete manipulation of the PDF by an end user. Mills will contact Beard to finalize faculty submission PDF generation settings (PDF/A-1b conformity, color model, target resolution, compression, etc.) and create a corresponding XML job ticket for the PDF server to use.

- 2) Collection statistics update, collection stats audit and load issue on mss3; Geng, Mills, Triggs.

Since the last meeting it was found that "Full Summary" statistics were causing the production system CPU usage to exceed acceptable levels and caused unacceptable response times from the Statistics software and other applications as well. A dump of the production database was placed on the development machine, Geng and Triggs examined it and saw somewhat better response times, however it was noted the applications performance is still unacceptable. The statistics package needs re-examination, the logic used to first construct the package is not scaling to the growth of the repository at all. For R4.5 the "Full

Summary” links will be removed from the statistics application so it is not accidentally run. The “Top 25” feature will now be limited by year, not the whole dataset. If a “Full Summary” is needed then a dump from the database will be placed on the development server and it will be run from there as not to disturb other production applications. Geng will continue to audit the statistics to find accurate ways of measuring object counts etc. Hoover mentioned that since the “Full Summary” bug was found it had been run on the production system. He will distribute the IP addresses of the machines that ran it to ensure this feature is not run on the production system.

3) Continue Google indexing discussion - we need PID on the ‘showfed’ display

A note will be added to the test Google objects on lefty that state they are truly coming from a “Test Google Index.” Showfed will display the handle ID of the object with the label “Persistent URL”, which is the same label that will be used in the search objects full record display. Discussion began about using “rucore” instead of “mss3” in the object URI. Virtual hostnames will be created for NJDH and RUcore on the development system.

4) Checking of handles on MSS3 report, audit of handles; Triggs.

Handles will be checked and if errors occur the script will generate a new handle. The script will be run on lefty, then mss2 and finally mss3. January 10th, 2008 is the expected due date for all environments to be corrected.

5) XACML follow up; Ananthan and Triggs.

Next Meeting.

6) Discussion with Fedora architect and impact on our release plan

Next Meeting.

7) Periodic audit of signatures on mss3 - are we doing this and who gets alerts?

Hoover ran the signature checking script on mss3, 12 hours runtime, and found 80 errors. 76 have the DS1, DS2... errors from last year still, 4 have multiple ARCH1 errors. Hoover will share the errors with Ananthan and Triggs. Ananthan and Triggs will run the signature checker on the lefty and either fix or delete the bad objects on lefty. Once this cleanup is finished the same will be done on mss2 with the assistance of Hoover. Finally mss3 will be cleansed of objects reporting bad signatures. This will all be done to establish a base on all systems where there are no bad signatures being found. ARCH1 checking should be more intelligent and not report collection objects and disseminators as errors. This should reduce amount “ARCH1 not found” errors due to the reporting of objects that are not supposed to have an ARCH1 to being with. The mellon signature file can be removed, it is not being used anymore.

Next Meeting, Thursday, January 10th, 2008

- XACML follow up; Ananthan and Triggs.
- Discussion with Fedora architect and impact on our release plan