

Software Architecture Working Group Meeting Minutes - April 21st, 2016

Attendees

Kalaivani Ananthan (recorder), Isaiah Beard, Jie Geng, Dave Hoover, Ron Jantz, Rhonda Marker, Chad Mills (chair), Jeffery Triggs, Yang Yu

Agenda

- 1) Announcements –All
- 2) R8.0 final issues (if any) – All
- 3) ABBYY OCR output as replacement to search XML datastream (XML1) – Mills, Yu
- 4) ABBYY licensing and software updates – Ananthan, Beard, Yu
- 5) R8.1 schedule – All
- 6) TEI text project – Jantz
- 7) RULMAIL address transition and impact on email notifications sent – All
- 8) Jobs & Reports – Triggs

1) Announcements

- a. Jantz
 - Fedora DC users' conference will be on April 27th and 28th. Yang is interested in attending, if time permits. Mike Durbin at UVA is presenting a session on using Fedora 4.0 on production. University of Maryland is presenting a second session on Moving to Fedora 4.0. Second day is dedicated to Hydra.
- b. Ananthan
 - Twenty JIRA licenses have been obtained from OIT. Ananthan will send the URL to the group to start using when time permits.

2) R8.0 final issues

- a. Announcements and notifications done. Hoover will start the pre-work ahead of time without touching the production system. Installation will happen between 7 am and 10 am.
- b. IP restrictions will be in place.
- c. Developers will be in TSB for testing at 10 am.

3) ABBYY OCR output as replacement to search XML datastream (XML1)

- a. **ABBYY OCR**
 - The XML (with coordinates) generated by ABBYY looks different than the one currently used in RUcore. ABBYY creates texts with blockType=Text, and in addition to attributes b, r, t, and l, there is a baseline attribute, probably indicating original "line" considered by ABBYY. As a result, highlighting in the interface would not be as accurate.
 - Chad will make changes to the front end code to detect the type of the OCR XML (DjVuXML or ABBYY XML).
 - ABBYY XML with coordinates will be stored as OCR-1.
 - Starting R8.1, we'll discontinue using XML-1; instead we will create OCR-1 datastream only!

- This will eliminate using the desktop version of ABBYY client and opens up for others to use this pipeline.

b. Open source software to convert URL/HTML to PDF

- ADLIB PDF server creates better quality than the open source software.
- Open source software is available on the rucore-dev server.
- Yang will explore options to configure the software to see if it can produce a better quality.

4) ABBYY licensing and software updates

- a. We've a new Account Manager. Yang, Nick, and Kalaivani had a WebEX session with technical support.
- b. The current ABBYY server needs to be upgraded to support MS Office documents.
- c. We're waiting to get the download link.

5) R8.1 schedule

- a. Code freeze date is Wednesday, April 27, 2016.
- b. Jeffery will enter a new request for dlr/EDIT to make the collections table editable by administrators.
- c. Test plan nearly complete. Kalaivani will send out a draft plan by Monday noon.
- d. Developers should review entries in their projects and update the status appropriately.
- e. If ABBYY with MS Office support software is not installed and functioning by Monday, we'll continue to use ADLIB PDF server in this release.

6) TEI text project

- a. Ron and Jim Niessen are working on a new project, which may result as a new collection in RUcore. RUL has a license in perpetuity to distribute these materials through RUcore.
- b. Approximately 160 volumes; TEI marked up XML text done offline; ready to go.
- c. Grace approved Ron to fill out the 4-part project initiation form.
- d. Materials will be restricted for Rutgers NetID and also by IP address.
- e. No special software needs; just configuration changes primarily to support this project.

7) RULMAIL address transition and impact on email notifications sent

- a. RULMail will be migrated to Rutgers Connect (Office 365); no firm date yet but we expect it to happen in June to August time frame.
- b. As part of migration, distribution lists will need to be recreated in Rutgers Connect.
- c. Developers should take an inventory of what email addresses are used in their code. If this is part of the code, it should be updated to configuration.
- d. Ananthan has sent an inquiry to OIT and will send us an update.

8) Jobs and Reports Update

- a. RARCH jobs suspended due to backup issues.
- b. Isaiah will delete RARCHs after verifying.
- c. Script currently does not compare checksum between RARCH1 and ARCH1. Done manually.
- d. One checksum mismatch uncovered. Isaiah will investigate.
- e. RARCHs should be deleted before running the next batch.
- f. Jeffery will review outstanding entries in software.libraries and close them as appropriate.