Implementation and Development Call 20130122

Chin Hua's questions before starting on the improvements to visualization planning:

Background

An exception was being thrown from VIVO 1.5.1. after at the University of Florida after their update from 1.4, during the process to recompute the search index. The VIVO code was throwing only a very obscure exception about an unexpected character that proved to be a form feed character that had been pasted from a PDF document via interactive editing.

Florida has now fixed the problem in their data that triggered the exception, but it would be helpful to know more about where the error happened so that VIVO can trap for this error and either just skip the record being indexed and identify which record has the problem.

Meanwhile, the visualization development team at Indiana has been trying to load the Florida data from a database dump to use as a test for debugging issues with very large visualizations.

How much memory was used at UFL to generate the index successfully?

- Chin Hua is working from a database dump from UF – that has to be restored into his local MySQL database at Indiana
- He had had problems even loading the ontology – are there customizations in the UF ontology?
    - They had sent him an ontology file beforehand as an N3 file – ask them for the directory it came from
        - What is the right step for adding an additional ontology?
        - Put the N3 file either in the productmods directory in the source tree or in the Tomcat webapps directory
            - look for WEB-INF/filegraph/tbox and the application will find it a startup time
            - VIVO will make sure that file has been loaded into the database
            - VIVO will also look to see whether there are classes and properties in the database no longer found in files in the WEB-INF/filegraph/tbox directory, and remove them from the database
- Shouldn't need much memory at all to build the index – it has to load all the URI strings – a couple hundred megabytes should be the max it will ever need
- Chin Hua needs to get a new copy of the UF VIVO data with the bad data record
- How long should it take to start up the VIVO? Not that long
- How long should it take to reindex the VIVO data from Florida after it has been loaded in a test instance at Indiana? Cornell can reindex in about 4 hours, but Florida's database is bigger
- You used to be able to tell how far the indexing has gotten, but with the 1.5.1 release couldn't see any progress information in the VIVO interface or in the logs (vivo.all.log) – if not seeing it, the indexing must not be happening – it had stopped in November because it had thrown an error
- The remaining question for the code is whether we need to trap for the form feed character that caused the error

What version of the code he should use?

- we recommend either the release files or the development branch on Github – if are planning on making changes, you will want to be working from Github
- the code in the development branch has installation documentation describing recent changes

What is the plan for visualization caching?

- looking at doing it in memory
- thinking of caching the results rather than the model
- need to load the UF data and see what is slow or fails before make any definite plans
- The VIVO log file page in the wiki describes how to control the logging level in a running VIVO instance to see more messages from any part of the code