

2020-05-08 Cornell LD4P2 Meeting notes

Date: 08 May 2020

Attendees: Lynette, Tim, Steven, Jason, John, Huda, Simeon

Regrets:

Agenda & Notes

Review actions from [2020-05-01 Cornell LD4P2 Meeting notes](#)

- [E. Lynette Rayle](#) QA performance
 - 2020-04-17 Did an analysis sorting response time by complexity and size but it didn't show a clear picture, will try a little more analysis next week. When Ligatus and CERL were added - why is Ligatus fast and CERL slow? Dave expects to have more time soon
 - 2020-04-24 Lynette did an analysis of performance to try to understand whether speed is clearly related to data size or complexity of extended context. Result are that there isn't clear correlation. Tried to parallelize parts of QA and in some places saw slowdown, one place found speed improvement where the complexity is high. However, in the complex cases the times are often still rather long (0.5–2s) but not markedly longer than somewhat less complex queries. Still the worst cases are because of the retrieve time from Dave's cache, he is looking at why CERL is slow when we might not expect it to be. Unfortunately no clear path to improving everything from the QA side: Lynette will try to understand why the OCLCFAST graph load is so slow.
 - 2020-05-08 No progress (Lynette working on exhibits about half the work, but has worked on accuracy), no updates from Dave
- [E. Lynette Rayle](#) to set up best practices working group around linked data APIs for authorities documenting on [Charter 1 - Best Practices for Authoritative Data Working Group](#)
 - 2020-05-08 Now starting first Monday in June and then every other week for 4 months, getting folks to do some work up-front. Think that a later WG might look at change management
- [Huda Khan](#) reflections on Knowledge Graph Conference – see slack comments. Meeting overall was very industry/enterprise including real-estate etc.. Good presentation by Oracle with information about a DB where they can do both SQL and SPARQL (e.g. start with relational DB, create RDF dataset from it, query either), many other presentations just assumed SPARQL or GraphQL. There was discussion of what is knowledge graph is, consensus that it is just a bunch of RDF or property graphs for knowledge/semantic representation. Not much discussion of performance, perhaps because much of the work is about offline analysis and then machine learning etc. Discussion from Yahoo! about rich cards (essentially knowledge panels)
- Blacklight summit happening May 7,8 virtual - [Huda Khan](#) presented demo video yesterday, expecting Jenn, Melissa, Frances to attend. Mostly demos yesterday. HathiTrust ETAS is mentioned a lot

Status updates and planning

- Enhanced Discovery - **WHAM!** (see also <https://wiki.duraspace.org/x/sJl7Bg> and <https://github.com/LD4P/discovery/projects/1>)
 - See: [Organizing doc](#) and [Pseudonym](#) thoughts
 - John has been working on a gem using overlapping namespaces as the Duke gem does
 - Tim is on D&A sprint
 - Had meeting on Monday to review use cases and scenarios with Jenn, Frances, Laura – see notes added to google doc.
- Authority Lookups for Sinopia (Lookup infrastructure: https://github.com/LD4P/qa_server/projects/2, Authority requests: https://github.com/LD4P/qa_server/projects/1)
 - 2020-05-06 discussion where Astrid shared feedback about Sinopia and QA from catalogers at <https://docs.google.com/document/d/14Sh2mBqkB2i9xml-Y7Aw-BGyvSAGwIS0l40jQXz88Pw/> . We note trade-offs between cached access and direct access in control /speed/scalability
 - 2020-05-08 Lynette has done work on merging in accuracy tests that Steven had produced. Everything is in the system pending deployment. Going to put the test harness in rspec so that it can be run in the background on stage. As much as possible will try to run the same tests on direct connection and on cache so we can do comparison. Think this should be straightforward
- Meetings (see [LD4P2 Cornell Meeting Attendances](#))
 - [LD4 Conference 2020](#), was to be May 13/14, 2020 but no dates set of virtual form
 - Planning going ahead for meeting to work out virtual presentations. There was a survey of presenters to understand how they might want to present. Steven likes the idea of recorded sessions followed by office hours.
 - 2020-05-08 Planning committee met, noted that presenters need a reasonable amount of time to prepare. The plan is to group sessions into panels and then reach out to speakers about scheduling. Content will be spread over several days. Leaning toward live sessions that will be recorded.
 - rdfs:seeAlso [Conferences Related to Linked Data in Libraries](#)
- Next meetings
 - ...