2020-02-21 Cornell LD4P2 Meeting notes

Date: 21 Feb 2020

Attendees: Lynette, Jason, Tim, Simeon

Regrets: Steven, John, Huda (though she brought cup-cakes)

Agenda & Notes

Review actions from 2020-02-14 Cornell LD4P2 Meeting notes

- Huda Khan to discuss with Astrid and David possible collaboration with U Chicago over usability (and maybe others in DOG team)
    - Notes: https://docs.google.com/document/d/1_V7JfAKqSn63-G0zupLZigt8o7R3v_WXC7hWlt8w744/edit
    - 2020-01-10 Chicago report will be discussed in DAG meeting on Jan 21 (expect to happen as scheduled). Follow up with David and Astrid to look at what we could apply from what they learned and what additional user studies they could help us with
    - DAG presentation did occur and it was very interesting (link to meeting) – in depth interviews to understand research needs with open-ended questions, need more thought to understand how we might apply lessons from this
    - 2020-02-14 Chicago uses vufind which already has a notion of similar items, questions of what types of semantic similarity might be interesting. Also discussion of author browse and articles; presentation of introductory materials for adjacent topics – they are going to discuss internally what to investigate further. Notes from U Chicago
- E. Lynette Rayle QA performance
    - 2020-02-14 Dave has made progress this week, is moving to have all data in index tailored to search in order to avoid SPARQL queries at search time, results are in a blob of RDF. Working on CERL first, expect to get this our soon. Will then try MeSH and OCLC FAST, then LC.
    - 2020-02-21 CERL was deployed with the new index strategy but no before and after to compare. However, this is small so we need to wait for LC or such to get a sense of possible improvement
- Adam Smith to investigate cost and any issues with setting up a D&A Beta system to allow broader testing of some discovery ideas from this work
- John Skiles Skinner to continue discussion with Hathi trust about an API or access to their index
    - 2020-01-31 There is investigation but not sure whether it will result in something we can use
    - 2020-02-14 Some more discussions with Hathi and suggestions was to use current search with debug facility that includes things like facet values in machine readable form (requires either 1) a user account for testing, or 2) to use IP access for our dev machine but there is some issue of fixed external IP for our dev VMs
    - 2020-02-21 HathiTrust sent over the XML version for one of the queries that John had tried for zero results. This would be the same xml they may be able to open up for us by allowing the IP address of my dev vm to access the URL that would result in XML. Huda set up a controller that parses the xml and returns json with the list of subject heading strings and set of search results being returned (this is the xml version of the search results page that includes subject facet values). John said he should be able to incorporate the subjects and perhaps the results into the zero search results page.
- Huda Khan Tim Worrall John Skiles Skinner to finish up lessons learned from BAM!
    - 2020-02-14 Will be done by next week...
    - 2020-02-21 progress on the BAM lessons learned doc and will aim for finishing that today...
- See items under travel/conferences

Status updates and planning

- Discovery presentation on 3/3: what is agenda? who is speaking? who is announcing to CUL?
    - Goal is to disseminate our week and then get some feedback about what might be promising
    - We need to have a way to take good notes – designate a note taker
    - View from staff involved with virtual libraries would be interesting, also D&A user reps
    - Need to be careful about managing expectations for stage this work is at and what might happen or not going forward
    - ACTION - Jason to write up, circulate and then send out on Tuesday – was DONE
    - ACTION - Huda Khan to collect everyone involved next week week to plan the session, review on 2/28
- Cataloging Sinatra and other 45's (Discogs data, https://github.com/ld4p/qa_server/issues?q=is%3Aissue+is%3Aopen+label%3ADiscogs)
    - Lookups for place not usable and hence places are not being recorded, relies on work from Dave to fix: https://github.com/LD4P/qa_server/issues/248 & https://github.com/LD4P/qa_server/issues/240
    - Have a currently insurmountable issue with nested profiles. When create Work profile with nested Instance profile there isn't a URI for the Instance (it just gets hung from a bnode). Without a URI the title of the Instance doesn't get indexed. The Sinopia team are unable to fix this in the near term.
    - Cataloging work continues with the above limitations
- Enhanced Discovery (see also https://wiki.duraspace.org/x/sJI7Bg and https://github.com/LD4P/discovery/projects/1)
    - SMASH! (dev to run through 7 Feb, then user testing, video and write-up)
    - Open meeting March 3, 2-3:30pm in Mann 102 and should Zoom it too
    - How will we decide what to take forward from KAPOW!, BAM! and SMASH!? (or as Tim put it, "what happens in late February?")
        - ACTION Everyone to think about what to take forward, plan to focus 2/28 meeting on this
    - SMASH!
        - 2020-02-21 We have a smash video demo time set up for Monday. Hitchcock homage and cameos will be done separately later in the week.
- Authority Lookups for Sinopia (Lookup infrastructure: https://github.com/LD4P/qa_server/projects/2, Authority requests: https://github.com/LD4P/qa_server/projects/1)
    - Current Beanstalk deployment on AWS is broken. The Ruby bundler is now not working for unknown reasons – cannot deploy with known good code state so it seems like something in the environment
    - When deployment issue solved... will then put out CERL and ligatus and extended context for MeSH along with sub-authorities. Also some refactoring associated with monitoring status page (including fixing a memory leak due to a long-used hash - ruby doesn't reclaim space from deleted entries)
    - Not sure whether Dave has redone the index to avoid SPARQL for MeSH – if it is done then we will have a comparison

- ACTION - Lynette to ask Tiziana about SHARE-VDE APIs for real-time up-to-date search and for possible engagement in linked data best practices for authoritative data working group
- Travel and meetings (see LD4P2 Cornell Meeting Attendances)
    - Knowledge Graph Conference (Columbia University), May 6-7, workshops 5/4-5/5
        - Workshops due 1/15 (not doing this), proposals due 2/28, talks 20mins
        - Huda Khan will lead effort to consider proposal
            - Preliminary brainstorming
    - LD4P2 cohort and partner meeting, LC, 20 & 21 April
        - Hotels are expensive those days
        - Partner meeting on 21 April - https://docs.google.com/document/d/17O58rsF_-N8n9hiGDYuxggdSRS0EXSPBiw2XgZTb4jw/edit#heading=h.jiyum6rb9zes
        - ACTION - Simeon to ask about timing of the meeting, wonder whether 3:30pm end would be possible?
    - LD4 Conference 2020 at College Station, TX (TAMU) - May 13/14, 2020
        - Proposals went in 2020-01-31 – Discovery, QA/Sinopia, BOF for best practices, Intro to LD workshop
        - Expect hear back soon, registration will open then (first come first served)
    - rdfs:seeAlso Conferences Related to Linked Data in Libraries
- Next meetings:
    - 2020-02-28 Lynette out