Request for Proposals

Workset Creation for Scholarly Analysis: Prototyping Projects

The Board of Trustees of the University of Illinois (hereinafter "University") has received an Award ("Prime Award") from the Andrew W. Mellon Foundation ("Foundation") in support of a project entitled, “Workset Creation For Scholarly Analysis: Prototyping Project” (WCSA). More information about the Prime Award project is available from http://worksets.htrc.illinois.edu. The overall goal of this project is to develop new tools and techniques to assist researchers and scholars in identifying and selecting resources from within the HathiTrust and creating worksets of these resources useful for their scholarly analyses.

RFP Schedule:

RFP Available: 22 November 2013
Letters of Intent Due (preferred): 16 December 2013
Final Proposals Due: 13 January 2014
Shortlist Meeting Invitations Issued: 20 January 2014
Shortlist Meeting: 20 February 2014
Award Notification: No later than 15 March 2014

Funding Available:

Proposals are solicited for prototyping projects to define and implement a tool or service that will help scholars better identify and select relevant resources at scale from the HathiTrust corpus and/or facilitate the construction of large-scale worksets useful for scholarly analyses; each prototype tool or service will be demonstrated over a representative sample (~250,000 texts) of the HathiTrust corpus provided by the Prime Awardee. Grants of $40,000 will be offered to each of 4 successful Respondents. Funds are to be expended over a period of 9 months beginning 15 April 2014. Funds are to be spent primarily on project staff salaries and benefits. Respondents are discouraged from seeking funds for computer equipment or software. A shortlist of 8 Respondents will be invited to send 1 team member to a workshop planned for February 2014 to present and elaborate upon proposed projects prior to final award selection. The Prime Awardee will reimburse reasonable travel expenses to this workshop for one person per proposal. Each of the 4 Respondents receiving an award will be required to submit a final report and to send 1 team member to a developer showcase in January 2015. Award funds may be expended to reimburse reasonable travel expenses incurred by a project representative to attend the developer showcase; requests for additional travel funding are discouraged. It is a condition of the Prime Award that funding not be spent on student tuitions, overhead, indirect costs (e.g., ICR), or the equivalent.

Program Description:

The HathiTrust (HT) is a large digitized-text corpus (> 10 million volumes) of keen interest to researchers working in a wide range of scholarly disciplines. To tap the analytic potential of this large and diverse corpus, to tame it and make it useful to them, many researchers need the wherewithal to gather together, into a kind of personal digital carrel, cohesive and coherent subsets of HT texts (potentially tens or hundreds of thousands of volumes or parts of volumes) amenable to the in depth forms of analysis they want to do. The attributes on which they seek to collocate digitized texts are not always recorded in standard bibliographic descriptions. This is illustrated by a sampling of questions heard from scholars attending the 2012 & 2013 HTRC UnCamps.

- "What materials does HT contain that pertain to Japan? How many volumes are in Japanese?"
- "Has anyone already built a definitive set of works that analyze the oeuvre of such authors as Dickens or Shakespeare?"
- "What musical scores are in the corpus? What works contain music notation?"
- "How would I gather works by 16th-century women? By 19th-century men?"
- "Which works are fiction? Which are non-fiction? Which are commentaries? Essays? Poetry? Prose?"
- "How would I gather together all volumes having images of Victorian England?"
- "Which versions of multi-copy or multi-edition works should I use in my experiment?"
- "How do I merge a HT sub-collection of works and metadata with my set of works and tags and my colleague's annotations?"

The Prime Award includes funding for 4 sub-awards to collaborate with the HTRC in conducting prototyping projects to develop and validate the potential of specific algorithms, services and/or tools that can facilitate the kinds of research investigations illustrated above. We are seeking proposals for small prototype experiments from engaged teams of digital humanists, librarians and computer scientists. We anticipate that the proposals received will approach the problem in a variety of ways and will undertake to address one or more of the following key research questions:

- Can we enrich the HathiTrust corpus metadata by distilling analytics over full text?
- Can we augment string-based metadata with URIs for recognized entities – e.g., names, subjects, publication location, etc. -- and by doing so can we leverage external services to facilitate discovery and clustering of resources?
- Can we leverage existing, well-defined external corpora to identify complementary subsets of HT volumes, and having done so can we demonstrate the ability to create and perform analytics over an integrated workset that includes resources external to HT?

Context: The HathiTrust Research Center

The HathiTrust Research Center (HTRC) is a collaborative research center launched jointly by Indiana University and the University of Illinois, along with the HathiTrust Digital Library, to help meet the technical challenges that researchers face when dealing with massive amounts of digital text. The HTRC is focused on developing cutting-edge software tools, services and cyberinfrastructure to enable advanced computational access to the growing digital record of human knowledge. Leveraging data storage and computational infrastructure at Indiana University and the University of Illinois at Urbana-Champaign, the HTRC is provisioning a secure computational and data environment for scholars to

perform research using the HathiTrust corpus. The center is breaking new ground in the areas of text mining and non-consumptive research. This will allow scholars to fully utilize content of the HathiTrust Library while preventing intellectual property misuse within the confines of current U.S. copyright law.

Eligibility:

Grants awarded through this RFP will be made to the institution, not to the individual project director. Participants from the University of Illinois and Indiana University cannot serve as lead PI and cannot be funded on this award, nor can staff employed directly by the HathiTrust. Not-for-profit research libraries, digital humanities centers, institutions of higher learning, and other 501(c) entities involved in the preservation and/or dissemination of cultural heritage or scholarly humanities information resources are eligible to apply for and receive awards through this RFP. Eligible entities may partner with commercial entities when submitting a response to this RFP, but only if the non-profit entity is the lead institution, and only if the commercial entity agrees to be bound by the constraints of the terms and conditions outlined below, including those pertaining to intellectual property rights regarding results of the research conducted using award funds. Not-for-profit entities from outside the United States are welcome to apply.

Prerequisites:

Project proposals submitted in response to this RFP must describe limitations of the status quo and establish that adequate community and technical infrastructure are in place to support the proposed work. Specifically, viable proposals must describe:

1. The primary audience of scholarly users that the proposed tool or service will aid.
2. The collections and worksets already available to this audience (from the HathiTrust or elsewhere), and the inherent limitations of these collections and worksets.

Intellectual Property Terms and Conditions:

The Foundation and the HTRC are committed to making high-quality digital scholarly resources and digitally-based scholarly services as broadly available as possible for educational and cultural heritage scholarly purposes. Accordingly it is an intent of the Prime Award and all sub- awards made through this RFP that all modified and/or derivative software products, source codes, processes, tools, techniques, architecture, prototypes, and/or related documentation (collectively "Software") and all reports, evaluations, analyses and other documents, other than Software documentation, prepared in connection with or otherwise relating to the Project (collectively, "Documents") be used for the greatest possible educational benefit. It is a requirement therefore that Software developed over the course of the Prime Award and all sub- awards be made available on a royalty-free, open source basis, pursuant to an open source license located at www.opensource.org.

Pursuant to these aims, project proposals submitted in response to this RFP:

1. Must agree to make the Software available on a royalty-free, open source basis, pursuant to an open source license located at www.opensource.org.
2. Must incorporate in the Software only those digital products that are distributed and/or made available under an open source license that would allow the Respondent to distribute the Software under the open source license referenced above.

3. Must include the agreement of the Respondent's institution (and any partner institutions involved in the proposed project) to abide by the Intellectual Property Terms and Conditions attached hereto. Sub-awardee agreement is a requirement of the Prime Award and cannot be waived. Evidence of Respondent's agreement will be inclusion with the proposal submitted of a copy of the Intellectual Property Terms and Conditions attached below signed by a cognizant official of the Respondent's institution (and any partner institutions) with authority to execute this agreement.

Other Terms & Conditions:

Successful Respondents may be required to provide standard Institutional Representations and Certifications and abide by University of Illinois at Urbana-Champaign Special Provision Certifications, Clauses, and/or Regulations in effect at the time subaward is executed. For a copy of current applicable forms in use, please contact David W. Richardson at GCOAward@uillinois.edu.

In addition a formal Statement of Work including a listing of Outcomes / Deliverables and detailed Budget will be generated from the proposal submitted, agreed to by all parties, and incorporated as part of the executed sub-award contract between the University of Illinois at UC and each successful Respondent to this RFP.

Any sub-award resulting from a response to this RFP will be issued on a cost reimbursable basis.

Proposal Submissions:

For full consideration, an electronic copy of the complete proposal must be submitted by the 5:00 pm (Central Time Zone U.S.) on Monday, January 13, 2014, and must include the following elements:

1. Proposal Cover Sheet (template available http://worksets.htrc.illinois.edu/worksets/?page_id=20)
2. Transmittal cover letter on institutional letterhead from the proposed project PI or other appropriate officer of the Respondent institution.
3. Proposal narrative (maximum 7 pages). This narrative should begin with an introduction that summarizes and gives rationale for the proposed prototyping project, addressing how the proposed work will enhance the creation of worksets using the HathiTrust corpus and how it will demonstrate and advance the overall goals of the HTRC (see above). Be sure to clearly identify the personnel (by name and title) who will work with WCSA to implement the proposed prototype if funded. The narrative also should include descriptions of:
    - The scope and expected deliverables from tool building/development and any other technical work that will be done as part of the proposed prototyping project; *any plans to leverage pre-existing code must be described here and intellectual property issues addressed (see discussion of IP Terms and Conditions above)*
    - How scholars' requirements, and if applicable, scholars' feedback will be gathered
    - How the group proposing the work is well positioned to carry out this project
4. Schedule of completion.
5. List of deliverables, which should include at a minimum:
    - Final report
    - Well-documented source code

    - Sample output demonstrating enhancements for workset creation
6. Budget and budget justification (i.e., how the budgeted resources are needed to achieve project goals and objectives). Salary, benefits, and level-of-effort over the course of the proposed project should be called out in the budget for each proposed project participant. All non-salary budget requests must be clearly and fully detailed and must be defended in the budget justification. In-kind contributions are encouraged (but not required); any planned in-kind contributions should be described in this section.
7. Signed Intellectual Property Terms and Conditions agreement (see below).
8. Project participant bio-sketch or CV (2-page maximum) for PI and other senior project participants; these should demonstrate adequate and appropriate expertise and provide the review panel an indication of investigator track record on prior grants in this or similar domains.

For these proposals other supporting documentation (letters of support, etc.) will not normally be needed and are not expected. Should you want to include any supporting documentation as part of your proposal, please limit extent (10 pages maximum) and include all supporting documentation collated into a single PDF file.

Proposals should be submitted electronically as a single zip file containing all proposal elements to htrc.wcsa@gmail.com. Each proposal element (cover sheet, letter of transmittal, summary & rationale, narrative, budget & budget justification, IP terms & conditions, bio-sketches) should be a separate file within the zip file. Acceptable formats for proposal elements are PDF and MS Word (.doc or .docx). Within one week of electronic submission, one hard-copy of the proposal with original signatures should be mailed to:

Megan Finn Senseney
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 East Daniel Street
Champaign, Illinois 61820

Final Reporting Requirements:

PIs will be required to submit a final project report (PDF format preferred) detailing accomplishments and findings with 45 days of completion of the Project. These reports will be provided to the Foundation and made public on the HTRC Website.

Projects will provide links to and/or copies of all Software and Documents created during the course of this project. In lieu of maintaining these materials online for 5 years at Respondent's institution, if electronic copies are provided, The Graduate School of Library and Information Science at the University of Illinois will undertake to maintain these resources online and accessible for 5 years from project completion, in accord with Foundation requirements (see Intellectual Property Terms and Conditions, below).

Projects will also provide copies of or links to any publications generated in conjunction with this project.

Review Panel & Selection Process:

All proposals will be reviewed by the PIs. The PIs will develop a short list of at least 8 proposals, which will be reviewed and vetted by a subset of the project advisory board before PIs select awardees. Advisory board members will recuse themselves from this process should a proposal originating from their home institution be submitted. In January 2014, select Respondents will be invited to present proposals at an RFP shortlist meeting scheduled for February 20, 2014 in Chicago, Illinois. Respondents will be notified of the PIs' final decision by email in March 2014.

The successful respondent will demonstrate:

- A feasible and adaptable use case grounded in the needs of an identifiable scholar or scholarly community;
- A project team well qualified to conduct the work proposed;
- A viable 9-month workplan and schedule of completion; and
- A set of activities scoped within the range of HTRC's ability to provide technical support (e.g., datasets, compute cycles, etc.)

Proposal teams that include demonstrable collaboration with scholars will be favored. Preference will be given to individuals who are part of the primary audience of scholarly users.

Respondents are urged to contact htrc.wcsa@gmail.com, in advance of proposal submission to discuss eligibility, project details, prerequisites, and HTRC support with a member of the project team.