

Digital Humanities Research Awards

Google has so far digitized over 12 million books in over 300 languages -- a significant fraction of all books ever published. This collection, much of which was previously available only in university libraries, has helped many disciplines in the humanities. Because of this vast increase in digitized information, new avenues of literary research are now possible.

We also know more could be done to facilitate this research. Sometimes humanities research consists of amassing and curating a private data set, and writing or customizing tools specifically for that data set. While that might be the quickest way to answer a particular research question, it does little to help other researchers with similar questions. We want to make it easy for people to share not just results, but the tools and intermediate data upon which future research can build.

Toward these ends, Google is creating a collaborative research program to explore the digital humanities using the Google Books corpus. Disciplines of interest include (but are not limited to):

- Linguistics
- History
- Classics
- Literature
- Philosophy
- Sociology
- Archaeology
- Anthropology

Some example projects to give you an idea of what we're thinking about:

- Building software for tracking changes in language over time
- Building software for tagging and identifying concepts, structure, or entities in text (possibly tailored to a specific domain or language)
- Creating utilities to discover books and passages of interest to a particular discipline, with support for annotations and collaborative research
- Developing systems for crowdsourced corrections to book data (e.g., OCR text) and metadata
- Generating marked up freely usable datasets (e.g., part-of-speech tagging for little-known languages)
- The testing of a literary or historical hypothesis through innovative analysis of a book corpus
- Analysis of the generative or creative processes revealed in texts

These are one-time awards for up to US \$50,000. Google may choose to renew the award for another year following review of the research at the conclusion of the first year. Where appropriate, we expect award recipients to make their software, utilities, datasets, or similar results freely available to others to use.

We are requesting proposals in this area from select researchers and faculty members, and we would be delighted with your participation. We expect to make several awards under this program, and welcome proposals that include investigators from multiple organizations. Proposals that share resources or funding with other efforts are also welcome.

Google may offer help in some instances by providing relevant subsets of the Google Books corpus (subject to copyright and metadata licensing) or by hosting data for researchers. For instance, we anticipate being able to provide frequency lists of words categorized by language, publication date, country, and subject; and a limited number of scans and plain text from books in the public domain. If your research requires a specific data set, feel free to contact Jon Orwant (orwant@google.com) about availability.

You should include the following in your proposal:

- Proposal Title
- Principal Investigator (PI) full name, contact information (postal address, e-mail, phone), affiliation (university, school, college and/or department)
- Collaborators within and outside your organization
- Research objectives and expected results
- Benefit to the research community
- Data or services needed from Google, if any
- Funding needed

Please be concise but clear. As guidance, we'd prefer the body of the proposal to be five pages maximum.

You may submit your proposals through [this website](#).

- In addition to completing the online form, you will need to upload an application packet in PDF format that includes your proposal and a short CV from the PI.
- Please choose "Other" as the primary topic.
- Please specify Jon Orwant (orwant@google.com) as your Google technical sponsor.

The deadline for submissions is April 15, 2010. We expect to make final decisions by the end of May. Please direct any questions you may have about the process for these awards to Leslie Yeh Johnson (lesliey@google.com).

Best,

Jon and Leslie