Digital Humanities Research Awards - Europe

Google has so far digitized over 12 million books in 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 humanities and social science research are now possible. We also know more could be done to facilitate this research; it's important for people to share not just results which might be culled from curating and building customized tools for a specific data set, but to also share tools and intermediate data upon which future research can build.

Toward these ends, Google created a research program to explore the digital humanities using the Google Books corpus earlier this year. [Twelve awards](#) were given out to researchers asking questions on topics ranging from Victorian literature to the automated extraction of synonyms. While we were very pleased with the outcomes of the awards, we realize that they were somewhat limited in scope given the focus on research programs within the United States. As such, we're happy to announce a second round of awards, this time focused in Europe, with the intention of providing both monetary support and technical collaboration to researchers in the digital humanities.

Our research interests are the tremendous advances now possible with the massive number of readily available digital texts (books, newspapers, blogs, etc.). Computational approaches to analysis across the combined bodies of texts are especially complementary to Google's digitization efforts and data-intensive approach to technology. In addition, as mentioned above, proposals for tools and datasets that would advance this type of research would also be in scope for this program.

Humanities and social sciences of interest include (but are not limited to):
- Linguistics
- History
- Classics
- Literature
- Philosophy
- Sociology
- Archaeology
- Anthropology

Some concrete examples of projects that might be of interest:

- 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

We are requesting proposals in this area from selected researchers and faculty members, and we would be delighted in your participation. We expect to make several awards through this program, and welcome proposals that include investigators from multiple organizations - especially those collaborations which address both the social science and computational aspects of this field. 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 Leslie Yeh Johnson (lesliey@google.com) about availability.

These are one-time awards for US $50,000, to which overhead costs cannot be applied. 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 strongly prefer award recipients to make their software, utilities, datasets, or similar results freely available to others to use under an open source Apache or BSD-style license.

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)
- Principal Investigator CV
- Collaborators within and outside your organization
- Please address the following in the proposal body
  - Research objectives and expected results
  - Benefit to the research community
  - Data or services needed from Google, if any (be aware that not all data will be automatically available)
  - Validity of approach (e.g. availability of data, computational soundess)
  - Description of funding usage

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 directly to me, Leslie Yeh Johnson (lesliey@google.com). One pdf file is preferred.

The deadline for submissions is November 1, 2010. We expect to make final decisions by the end of November. Please direct any questions you may have about the process for these awards again to Leslie Yeh Johnson (lesliey@google.com).

Best,
Leslie Yeh Johnson