Brandeis University Receives a Major Grant to Enhance the LAPPS Grid Project and EU's CLARIN Platform to provide access to NLP-enabled tools to quickly analyze huge amounts of language for digital humanities to create Smart Archives.

Brandeis announced today that it is has been awarded a 16-month $673,000 grant from the Andrew W. Mellon Foundation to expand and deploy the Language Application (LAPPS) Grid Project that connects open-source computer programs to analyze texts from diverse sources and genres. The programs analyze any language content, determine the overall meaning, and help uncover hidden relationships embedded in the data. According to James Pustejovsky, the Project Director, the Mellon Foundation support will allow Brandeis and its collaborators from around the world to extend the range of the LAPPS Grid platform by linking it to a similarly broad and extensive one known as the European Common Language Resources and Technology Infrastructure (CLARIN). Dr. Pustejovsky explains that, "while our previous Mellon grant focused on the design of an 'internet of language applications', the current grant will allow us to actually implement the architecture for the everyday computer user." The project team, which includes Brandeis University, Vassar College, University of Tübingen, and Charles University in Prague, plan to bring these two big services together. In addition, the project will demonstrate the use of a suite of tools accessible from both platforms to create rich metadata sets for text archives, allowing users to explore digital data on both sides by zeroing in on semantically meaningful expressions in the texts, for example, names, products, times, events, companies, countries, geo-locations and relations. We will apply these metadata enrichment algorithms over two archives, The American Archive of Public Broadcasting (AAPB) and the German and Czech Parliamentary Proceedings (GCPP).

The LAPPS Grid is an NSF-funded collaborative effort among US partners Brandeis University, Vassar College, Carnegie-Mellon University, and the Linguistic Data Consortium at the University of Pennsylvania. The effort resulted in a terminology for a core of linguistic objects and features exchanged among NLP tools that consume and produce linguistically annotated data, as well as in protocols to tie together a variety of language processing services in processing pipeline requested by a user. The new project will continue the work performed under the previous Mellon grant that funded the team to explore and plan the seamless integration of the two platforms.

"We believe that this framework has the potential to transform scholarship and development across multiple disciplines in the sciences, language and social sciences, and digital humanities," says Pustejovsky, the TJX Feldberg Chair of Computer Science at Brandeis, "by providing a transparent interface to a massive range of tools and resources, at a level unprecedented in the past."

The other lead investigators in on the Mellon grant are: Nancy Ide (Vassar College), Erhard Hinrichs (University of Tübingen) and Jan Hajic (Charles University in Prague).

For more information regarding the scope of this LAPPS GRID project, or to speak with someone at Brandeis please contact me at your convenience at jamesp@brandeis.edu, or 781.736.2709.