

textgrid.de

Article Torsten Reimer - TextGrid

8-10 minutes

“Textgrid” project develops digital infrastructure to allow humanities researchers and scientists to collaborate globally on massive datasets

by Dr. Torsten Reimer, Programme Manager, Digital Infrastructure

Joint Information Systems Committee (JISC), London

Musty archives, dusty libraries – traditionally the natural habitat of humanities researchers, whilst scientists do their research in high-tech labs, or on supercomputers designed to simulate the world. Humanities researchers seem to be content with a book-covered study that looks as if a typewriter has only just lost its last stand against a laptop there. This may be the popular way to imagine how these disciplines work.

But now it's time to imagine new ways of working. Let's imagine that different humanities disciplines and librarians join forces to build their own digital tools, similar to the simulators of the climatologists or astronomers. Can we also imagine that

they reinvent their study rooms and libraries in virtual space, in order to share their research with colleagues all over the world? How about that they are using the latest technology to achieve this, technology that is usually reserved for particle physicists or computer scientists? Believe it or not, all of this is exactly what the German “TextGrid” project has been doing for the past few years.

“TextGrid” uses a technology that has been developed to answer some of the biggest questions of humankind: grid computing. Astronomers are using giant telescopes to find the origins of the universe in space; physicists use the Large Hadron Collider (LHC) to find them in Switzerland; and biologists decode the human genome on supercomputers. All of them create gigantic amounts of data. The LHC, for instance, generates 15 petabytes per year – that equals more than 7 trillion pages of text. Grid technology has been developed to store, analyse and share this gigantic amount of data with research groups all over the world. The technology allows researchers to connect computers and data and to run complex parallel analyses on distributed computer networks, no matter where their physical location might be.

“TextGrid” uses this architecture to enable researchers to jointly research our cultural heritage. The humanities now, too, needs a large-dataset solution, making “TextGrid” a timely innovation. The amount of data created, for instance, through text digitisation is now growing dramatically. Through

databases, we already have access hundreds of thousands of ancient inscriptions, medieval manuscripts and modern prints. Libraries and companies such as Google continue to digitise millions of books.

The potential of Internet technologies and digitisation of cultural heritage goes beyond giving us easy access to the world's archives, museums and libraries directly from our homes. Online projects such as Wikipedia demonstrate how the Internet can be used for collaboration across the globe. Researchers are increasingly using these technologies to jointly conduct new research in virtual space. The vision behind these Virtual Research Environments is to make it easy to give specialists access to the data and tools they need to solve problems. The German Research Foundation (DFG) and the German Federal Ministry of Education and Research (BMBF) fund such virtual environments to facilitate research collaboration across international and discipline boundaries.

In order to make this vision a reality, researchers, data and technological infrastructure have to be brought together. Funded by the BMBF, the pilot project "TextGrid" is working to achieve exactly that. Since 2006, ten German research institutions are working together to develop a Virtual Research Environment for the Humanities. The main focus is the distributed research on text, text ranging from 18th century encyclopaedia to rare manuscripts and digital music editions. Grid computing makes it possible to bring together this ever-

increasing amount of data and to give researchers the tools they need to collaboratively work on these materials.

Virtual Research Environments like “TextGrid” enable us to find new answers to old questions – or to even ask entirely new questions. The automatic comparison of thousands of texts, for instance, makes it much easier to find out where authors got their inspiration from or how ideas spread across the globe. Philologists can more systematically research how language has developed, and it becomes easier to go beyond elite culture and analyse the mass of popular writings. But “TextGrid” does much more than allowing individual researchers to search through large corpora of text. It also makes it possible to collaboratively work on and edit texts, opening up new possibilities for scholarly editions. According to Professor Fotis Jannidis from the University of Würzburg, visualisation, as it is supported by “TextGrid”, is a key technology in this context:

“With a click of the mouse, the reader can now decide which versions of a text they want to see in the digital environment. They can display annotations and text variants or select a dynamic display that puts different version right in front of their eyes. Visualisation is one of the big opportunities of digital editions.”

Digital editions allow experts from different countries to synchronously collaborate on decoding difficult text passages, while other colleagues work on teaching the meaning of texts

to computers. This makes it possible to, for example, perform a search over historical texts to find mentions of a place, no matter whether its current or historical place name is used. It also allows automatic linking of the name of a person to their biography, so that researchers can get further information on the person or verify who they actually are.

In an environment such as “TextGrid” data itself becomes a tool, as Professor Andrea Rapp from the Technical University Darmstadt explains:

“Working with a literary text can, for instance, lead to the question of whether a certain word is used specifically by this poet, and whether its use is regionally limited or very widespread. Interlinked dictionaries can give me answers to those questions. In this way data created by one researcher can become a new tool for another one. The prerequisite for this is a critical mass, though.”

“TextGrid” will not only make this critical mass available: it will also provide the environment in which such research questions can be answered. Findings can also be published within the tool and will be preserved for later use.

Making data, tools and research findings available online also makes it easier to bring together experts from other disciplines to answer tricky questions. Even today, physicists, librarians, computer scientists, linguists and other researchers work together to build Virtual Research Environments like “TextGrid”. In the future, this infrastructure will enable them to

collaboratively break new ground in answering research problems. This interdisciplinary working is one of the most important reasons why the DFG funds such projects, as programme director Dr Sigrun Eckelmann explains:

“Lately, we’ve started to see collaborative projects encompassing disciplines who did not previously interact much with each other, such as linguists and biologists or climatologists and historians. This is now possible and it constitutes a substantial qualitative change.”

Libraries play a key role in the development of these projects, as Dr Heike Neuroth from the University of Göttingen emphasises:

“Projects such as “TextGrid” allow us to better understand how research is conducted in particular disciplines so that we can best support them with tools and data. This is, of course, not a new mission for research libraries but has a tradition developed over millennia. In today’s knowledge and information society, this mission is supplemented by the new technologies. Providing and preserving books for research use will certainly continue to be a task for libraries.”

[Back to press kit ...](#)