

TextGrid

Virtuelle Arbeitsumgebung für die Geisteswissenschaften

Martina Kerzel, Jens Mittelbach, Thorsten Vitt

Dies ist ein Bericht über das Projekt TextGrid¹, das von Februar 2006 bis Mai 2009 durch das Bundesministerium für Bildung und Forschung² gefördert wurde. Das Ziel des Projekts ist die Entwicklung einer offenen Plattform, die vielfältige Bedürfnisse und Perspektiven im Gebiet der Textwissenschaften einbeziehen wird. Der Fokus liegt in erster Instanz auf den Textdaten und dem Bereich der Editionswissenschaften in funktionaler Verknüpfung mit den anderen Bereichen³. Weitere fachspezifische Funktionalitäten und zusätzliche Datentypen können aber aufgrund der Flexibilität der zugrunde liegenden Infrastruktur mit wenig zusätzlichem Aufwand integriert werden.

1 Einleitung

Neue Technologien verändern rasant die Wissenschaftslandschaft und die Arbeitskultur in den Geisteswissenschaften. Dies lässt sich insbesondere an der rasch wachsenden Zahl permanent verfügbarer digitaler Forschungsobjekte ablesen: von digitalisierten Quelltexten über digitale Film- und Tonaufnahmen bis hin zu Fachliteratur, Forschungsdaten und virtuellen Korpora.

Dank neuer computergestützter Methoden können solche Objekte in zunehmendem Maße erzeugt werden. Auch auf die Forschung selbst haben die direkte Verfügbarkeit digitaler Objekte und die vielfältigen Möglichkeiten, sie mit Hilfe von Computertechnologie auszuwerten, Auswirkungen. Forschungsfragen können heute dank neuer Technologien präziser formuliert und oft auf neue Weise beantwortet werden oder es können sogar gänzlich neue Fragestellungen entwickelt werden. Ein Blick auf die aus der technologischen Revolution resultierenden Notwendigkeiten und auf die gestiegenen Anforderungen an die Nutzer macht schnell die Herausforderung deutlich, vor der Wissenschaftler, Institutionen und Förderer in Deutschland zurzeit stehen: einerseits die Etablierung einer Forschungsinfrastruktur für die Geisteswissenschaften auf Basis von e-Science-Technologien – für die eHumanities – und andererseits die Entwicklung von virtuellen Forschungsumgebungen in Kooperation mit Wissenschaftlern, Bibliotheken, Rechenzentren und IT-Experten.

TextGrid entwickelt in einem ersten Schritt eine Arbeitsumgebung für die Geisteswissenschaften, in der Textressourcen kollaborativ ediert, annotiert, analysiert und ver-

öffentlicht werden können. Ziel ist, hier alle für den wissenschaftlichen Arbeitsablauf notwendigen Werkzeuge und Dienste bereitzustellen und einen weitgehend uneingeschränkten und einfachen Zugang zu den zahlreichen und stetig anwachsenden wissenschaftlichen Quellen bzw. Texten zu gewährleisten. Sicherheit bieten ein feingranulares Rechtemanagement sowie ein Repositorium für die langfristige Aufbewahrung der Daten.

Insgesamt spielen weniger die durch TextGrid beispielhaft bereitgestellten Werkzeuge eine Rolle (die, zum Teil in größerem Funktionsumfang, auch in lokalen Applikationen verfügbar sind), als vielmehr das Grid als eine bisher in den Geisteswissenschaften ungenutzt gebliebene infrastrukturelle Komponente. Die Nutzung dieser neuen Technologie macht ortsunabhängige, gemeinschaftliche Arbeit an wissenschaftlichen Fragestellungen zu Texten, umfangreichen Textsammlungen und Textkorpora und allem, was in Struktur und Umfeld damit zusammenhängt, möglich.

2 Das Grid in TextGrid

Ist das World Wide Web ein Netzwerk, das allen Teilnehmern dieses Netzwerkes Informationen zur Verfügung stellt und teilt, so ist das Grid ein weltumspannendes Netzwerk von Computern, die ihre Rechenleistung und Ressourcen dem gesamten Netzwerk zur Verfügung stellen⁴. Das vielleicht bekannteste Beispiel für das Grid-Computing ist die Verarbeitung der Datenmengen, die bei Versuchen mit dem Large Hadron Collider (LHC) am europäischen Forschungszentrum CERN⁵ anfallen. Hier sollen jährlich etwa 15 Petabyte Daten entstehen, auf die Wissenschaftler weltweit zugreifen können. Darüber hinaus werden diese Daten auch in einem DataGrid permanent gespeichert und verfügbar gehalten⁶.

Die Notwendigkeit, eine Forschungsinfrastruktur mit Grid-fähiger Architektur auch für die Geisteswissenschaften zu schaffen, liegt auf der Hand: Zum Einen können Grids ge-

¹ <http://www.textgrid.de/>

Das Projekt wird kollaborativ von universitären und kommerziellen Partnern durchgeführt. Beteiligt sind die Niedersächsische Staats- und Universitätsbibliothek, die Technische Universität Darmstadt, das Institut für Deutsche Sprache, das Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier, die Fachhochschule Worms, das Kompetenzzentrum für EDV-Philologie an der Universität Würzburg, die DAA-SI International GmbH und die Saphor GmbH.

² <http://www.bmbf.de/>

³ Die in einer weiteren Förderphase vorgesehenen Bereiche sind z.B. klassische Philologien, Musikwissenschaften, Psycholinguistik und Kunstgeschichte. Weitere Möglichkeiten sind in den *TextGrid-Szenarien* nachzulesen: http://www.textgrid.de/fileadmin/TextGrid/TextGrid-Szenarien_061212.pdf.

⁴ Dabei sorgt das Wort "grid" nicht ganz unbeabsichtigt zu einer Analogiebildung mit dem Stromnetz: Ähnlich einfach wie die Nutzung von Strom durch das Einführen eines Steckers in die Steckdose soll auch die Nutzung von Daten und Diensten innerhalb des Grid sein: <http://www.gridcafe.org/what-is-the-grid.html>.

⁵ <http://www.cern.de/>.

⁶ <http://lcg.web.cern.ch/LCG/public/default.htm>.