

TextGrid – Ein Serviceangebot für die kollaborative Arbeit an Personendaten
von Joachim Veit und Daniel Röwenstrunk

Das in den Forschungsverbund D-Grid (<http://www.d-grid.de/>) eingebettete Projekt TextGrid (<http://www.textgrid.de>) ist die erste Initiative zum Aufbau einer modularen Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung in den Geisteswissenschaften in Deutschland. Seit 2006 verfolgen die nunmehr zehn Partner aus den Geisteswissenschaften, Künsten, zusammen mit Bibliotheken und IT-Spezialisten das Ziel, ein Community-Grid aufzubauen, das kollaboratives Arbeiten der Fachwissenschaftler an verteilten Informationen, teils halb-automatische Analyse und Anreicherung der Daten, eine Onlineplattform für die Publikation derselben und eine garantierte Langzeitarchivierung ermöglichen soll.

Das TextGrid-Repository hält sämtliche im Grid verfügbaren Ressourcen wie zum Beispiel Wörterbücher, Lexika oder in Kürze vielleicht den gesamten Inhalt der größten deutschsprachigen Volltextbibliothek vor und bildet damit das Rückgrad des Systems. Über eine mögliche Integration kontrollierter Vokabulare wie der Personennormdatei ließe sich ein schneller ortsunabhängiger Zugriff auf diese Informationen realisieren, die gleichzeitig für Services rund um die automatische Erkennung und Auszeichnung von Personen in Texten genutzt werden könnten. Neben diesen „externen“ Daten legt das Repository die von den Nutzern erstellten Texte auf mehreren Servern im Grid verteilt ab, um so eine hohe Erreichbarkeit der Daten sowohl geographisch als auch in der Zeit zu gewährleisten. Im Kontext von Personendateien ließe sich in dieser Struktur eine Personendatenbank basierend auf XML-codierten Informationen aufbauen, die projektübergreifend, durch ein Rechtemanagement kontrolliert, bearbeitet und genutzt werden könnten.

Aber nicht nur das Nutzen, Anlegen und Verwalten personenbezogener Daten, auch die inhaltliche Verknüpfung dieser Personen ließe sich über eine bereits bestehende Triplett-Datenbank realisieren. Eine Kombination aus zwei Personen und einer Beziehungsbeschreibung würde die Modellierung etwa von Eltern-Kind- oder Lehrer-Schüler-Beziehungen ermöglichen.

Standardisierte Schnittstellen erlauben den Zugriff auf die Daten des Repository aus verschiedenen Anwendungen und Anwendungsszenarien heraus. So könnten Internet-basierte Darstellungen von Personen, Briefen oder anderen auf Texten und Bildern basierenden Informationen mittels dieser Schnittstellen und dank persistenter Identifikationsnummern die im Repository vorgehaltenen Daten visualisieren und einer breiteren Öffentlichkeit zugänglich machen. Eine Anwendung, die über den Einsatz dieser Schnittstellen die Daten des Repositories verfügbar macht, ist das im TextGrid-Projekt entwickelte TextGrid-Laboratory (kurz TextGridLab). Diese Arbeitsumgebung bietet die Möglichkeit, Projekte und Dateien anzulegen und kollaborativ zu bearbeiten. Über ein flexibles Rollenkonzept lassen sich Bearbeitungs- und Lese-Rechte dieser Objekte verwalten, sodass nicht nur auf mehrere Standorte verteilte Gruppen gemeinsam an den Daten arbeiten, sondern auch über die Projekte hinaus der Austausch und die Anreicherung von (Personen-)Daten feingranular gesteuert werden könnten. Zusätzlich bietet das TextGridLab die Möglichkeit sowohl den Einsatz als auch die Ergebnisse extern angebundener Webservices – wie zum Beispiel das in Entwicklung befindliche OCR von in Fraktur gesetzter Drucke – zu steuern.