

DHd-Tagung 2015

»Von Daten zu Erkenntnissen:

Digitale Geisteswissenschaften als Mittler
zwischen Information und Interpretation«

Workshop Proposal

Frosch oder Prinz?

TUSTEP als Werkzeug der Digitalen Geisteswissenschaften –
ein Workshop der
International TUSTEP User Group

Thomas Kollatz · Ute Recker-Hamm · Matthias Schneider

Frosch oder Prinz?

TUSTEP als Werkzeug der Digitalen Geisteswissenschaften – ein Workshop der International TUSTEP User Group

Das Tübinger System von Textverarbeitungsprogrammen (TUSTEP)¹ ist eines der ältesten, seit über 40 Jahren bis heute beständig weiterentwickelten und in zahlreichen geisteswissenschaftlichen Projekten² erfolgreich eingesetzten Programme der Digitalen Geisteswissenschaften. Es bietet einen Werkzeugkasten von speziell auf philologisch-sprachwissenschaftliche Anforderungen zugeschnittenen, auf einander abgestimmten Werkzeugen, die in hohem Maß an projektspezifische Erfordernisse angepasst werden können. Die Bandbreite der Programmmodule reicht dabei vom Vergleichen und Kollationieren, dem Zerlegen, Annotieren und Indexieren von Texten in XML oder anderen Formaten über das Skripting bis hin zum professionellen Satz und kann – beispielsweise für dynamische Webpublikationen – auch direkt auf Servern eingesetzt werden. Seit Kurzem verfügt TUSTEP über eine XML-basierte, alternative Kommandosyntax namens TXSTEP, die eine Steuerung des Programms von außerhalb, z.B. aus der XML-Entwicklungsumgebung Oxygen erlaubt.

Obwohl die Leistungsfähigkeit und hohe Qualität des Programmpakets national wie international unbestritten ist,³ eilt ihm auf Grund der reduzierten Benutzeroberfläche und seiner vermeintlich besonderen Arbeitsweise der Ruf voraus, schwer erlernbar zu sein. So kommt beispielsweise das TEI-Wiki – nachdem die Vorzüge von TUSTEP gelobt worden sind – zum Schluss »learning TUSTEP is not a walk in the park [...] a scholar with no or minor technical background will need support to get things work.«⁴

Diese Aufgabe übernimmt die International TUSTEP User Group (ITUG e.V.)⁵, die seit über 20 Jahren TUSTEP-Anwender und -Lerner durch Kurse, das TUSTEP-Wiki⁶ und

¹ tustep.uni-tuebingen.de

² Für Projekte in Auswahl, die TUSTEP einsetzen, siehe: itug.de/projekte.html

³ Siehe z. B. John Bradley: Text Tools, in: A Companion to Digital Humanities, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004 – digitalhumanities.org/companion/; Susan Hockey: The History of Humanities Computing, Ebenda; Michael Sperberg-McQueen: tustep.uni-tuebingen.de/sperberg2007.html; John Unsworth: new.livestream.com/accounts/2263400/events/3512861

⁴ wiki.tei-c.org/index.php/Publishing_printed_critical_editions_from_TEI

⁵ itug.de

⁶ tustep.wikispaces.com/TUSTEP-Wiki

Fachkolloquien unterstützt sowie die Weiterentwicklung von TUSTEP fördert, das seit mehreren Jahren als Open Source Software kostenlos verfügbar ist.

Der Workshop »Frosch oder Prinz?« der ITUG im Rahmen der DHd 2015 soll zeigen, dass es vielleicht nicht gerade ein Spaziergang im Park ist, TUSTEP zu erlernen, aber gewiß auch keine Hochgebirgstour. Dafür werden drei Programmmodule (Vergleich / Kollationierung, Skripting, Satz) exemplarisch heraus gegriffen und mit den Workshop-Teilnehmenden anhand eines durchgehenden Beispiels (»Froschkönig« der Brüder Grimm) praktisch erprobt. Der Workshop ist als Hands-on Workshop konzipiert.

VERGLEICHEN

Texte miteinander zu vergleichen, kann in verschiedenen Situationen nützlich sein: Neben dem üblichen Vergleichen und Kollationieren von Textvarianten spielen auch Qualitätssicherungsaspekte eine Rolle, wie z. B. das Feststellen von unerlaubten oder versehentlichen Eingriffen in Texte oder der Vergleich von mehrfach erfassten Eingabedaten zur Herstellung eines korrigierten Texts.

Für den Vergleich und die Weiterverarbeitung der Vergleichsergebnisse stellt TUSTEP mehrere Programmbausteine zur Verfügung. Das Vergleichsmodul erlaubt die Ausführung in verschiedenen Modi, die Spezifikation von als gleichwertig zu betrachtenden Zeichenfolgen, die Definition von Aufsatzpunkten u.v.a.m. Das Ergebnis des Vergleichs wird in Form einer Protokoll- bzw. Korrekturdatei ausgegeben, die manuell oder programmgesteuert weiter verarbeitet und / oder in die Ausgangsdaten rücküberführt werden kann. Der Vergleich von mehr als zwei Textzeugen ist dadurch möglich, dass jede Textvariante separat mit dem Grundtext verglichen wird und die Vergleichsprotokolle anschließend automatisch kumuliert werden.⁷

Im Workshop werden die Arbeitsweise des Vergleichsmoduls und die Möglichkeiten der Weiterverarbeitung der daraus resultierenden Daten anhand von zwei Auflagen des »Froschkönig« erprobt.

⁷ Dies Vorgehen impliziert keine vorweggenommene editorische Entscheidung für einen Archetyp, sondern stellt lediglich ein technisches Vorgehen dar, das in beliebiger Reihenfolge und beliebig oft in verschiedenen Abfolgen ausgeführt werden kann.

SKRIPTING MIT TUSCRIPT

Mit TUSCRIPT stellt TUSTEP eine vollwertige Skriptsprache bereit, die insbesondere zur philologischen Bearbeitung und Analyse von Textdaten geeignet ist.⁸ Im Workshop werden den Teilnehmenden einige Skripting-Funktionen sowie exemplarische Dateizugriffe mit TUSCRIPT am »Froschkönig« näher gebracht. So wird in einem prototypischen Arbeitsablauf ein Import des Textes aus einem Büroformat (MS-Word), sowie dessen Aufbereitung in eine TEI-XML konforme und valide Datei mit den Teilnehmenden durchgeführt. Anschliessend soll gezeigt werden, wie Texte mit einfachen Mitteln (semi-)automatisch angereichert und ausgezeichnet werden können. Die Skriptsprache eignet sich selbstverständlich auch zur systematischen Textanalyse und Recherche. Im Workshop werden zudem komplexe Suchfunktionen, Indexerstellung und dergleichen mehr vorgestellt.

SATZ

TUSTEP stellt zwei Programmbausteine für die Aufbereitung von Textdaten mit dem Ziel der Ausgabe als Postscript- respektive PDF-Datei bereit. Mit dem Modul #SATZ können einfache Texte bis hin zu anspruchsvollsten Editionen inklusive zeilensynoptischer Ausgabe sowie mehreren Apparaten und Registern hergestellt werden. Ein sogenannter »Roundtrip«, also das Zurückfließenlassen von Satzdaten (z. B. Seiten / Zeileninformationen) in die (XML-)Ausgangsdaten (z. B. zur nachträglichen Registererstellung) ist möglich. Das Modul *SATZ hingegen stellt einen einsteigerfreundlich komprimierten Ausschnitt dieses Funktionsumfangs zur Verfügung, welcher für die Produktion von Hausarbeiten, Dissertationen und Monographien oder Sammelbänden ausreichend ist. Hierfür stehen ein XML-Tagset, eine spezielle GUI, verschiedene Dokument- und Formatvorlagen sowie Schriftdefinitionen zur Verfügung. Die Lernkurve ist entsprechend flach. Das Modul *SATZ überzeugt nicht nur durch Benutzerfreundlichkeit, sondern darüber hinaus auch durch die vielfältigen Verwendungsmöglichkeiten out of the box (u. a. klassische Druckvorstufe für die Druckerei, on the fly-PDF-Erstellung aus Online-Datenbanken sowie die plattform- und medienunabhängige PDF-Erstellung).

⁸ Für Beispiele s. rosettacode.org/wiki/TUSCRIPT

Im Rahmen des Workshops soll das Werkzeug *SATZ benutzt werden, um die Ausgabe der zuvor mit den Vergleichs- und TUSCRIPT-Programmbausteinen aufbereiteten Textdaten als PS-/PDF-Datei zu exemplifizieren. Hierbei wird großer Wert auf die schnelle und komfortable Erschließung des Werkzeugs sowie die sehr guten Ergebnisse der Druckvorstufe gelegt.

Die Teilnehmenden erhalten die Möglichkeit, den in den vorhergehenden Abschnitten verarbeiteten Beispieltext selbst in unterschiedlichen Formaten zu erzeugen sowie mit der Druckausgabe zu experimentieren (Spaltensatz, Schriftarten, Bearbeitung des Textes in der *SATZ-GUI).

TUSTEP-Café

Während der Pause sowie unmittelbar vor und nach dem Workshop erhalten Projekte, die TUSTEP anwenden, Gelegenheit ihre Arbeit in informeller Atmosphäre durch Poster vorzustellen. Einige Zusagen liegen bereits vor. Falls der Workshop wie geplant durchgeführt werden kann, wird zudem ein »call for posters« über die ITUG Mailing Liste ausgerufen.