

Christian Thomas

Berlin-Brandenburgische Akademie der Wissenschaften

Deutsches Textarchiv, www.deutschestextarchiv.de

CLARIN-D, www.clarin-d.net, www.deutschestextarchiv.de/clarin_kupro

Humboldt-Universität zu Berlin

Hidden Kosmos: Reconstructing A. v. Humboldt's »Kosmos-Lectures«,

www.culture.hu-berlin.de/hidden-kosmos

Abstract zur Vortragsreihe Digital Humanities – Theorie und Methodik, Leipzig, 20.10.2015

Im Vortrag werden Primärquellen, die über das DFG-geförderte Projekt „Deutsches Textarchiv“ (DTA, www.deutschestextarchiv.de) der Berlin-Brandenburgischen Akademie der Wissenschaften verfügbar gemacht wurden, und die vielfältigen Möglichkeiten zu deren Nutzung präsentiert. Das DTA umfasst derzeit knapp 2800 Dokumente mit mehr als 630 000 digitalisierten Seiten und ca. 1,1 Mrd. Zeichen (Stand: 7.10.2015). Neben dem Anspruch, vielseitig nutzbare und qualitativ hochwertige Primärquellen frei verfügbar zu machen, liegt der Fokus des DTA-Projekts auf der korpus- bzw. computerlinguistischen Analyse der elektronischen Volltexte. Alle Quellen stehen frei über das Internet in verschiedenen Formaten zum Herunterladen und auch zum ‚Harvesten‘ über eine API bereit (www.deutschestextarchiv.de/api).

Ein Schwerpunkt des Vortrags liegt auf einer allgemeineren Einführung in die Grundlagen XML-basierter Annotation von Volltexten entsprechend den Empfehlungen der Text Encoding Initiative bzw. dem darauf basierenden DTA-Basisformat (DTABf, www.deutschestextarchiv.de/doku/basisformat). Das DTABf bietet inklusive einer ausführlichen, mit Beispielen illustrierten Dokumentation alle notwendigen Voraussetzungen, um einen heterogenen Dokumentenbestand strukturell und semantisch zu annotieren.

Anschließend an diesen einführenden Teil stehen die vielfältigen Nutzungsmöglichkeiten solchermaßen standardkonformer, plattformunabhängiger und (prinzipiell) interoperabler Textauszeichnung für die Publikation und computergestützte Analyse digitaler Editionen im Mittelpunkt. Anhand konkreter Beispiele aus der laufenden Kooperation des DTA mit dem Projekt „Hidden Kosmos: Reconstructing A. v. Humboldt's »Kosmos-Lectures«“ (<http://www.culture.hu-berlin.de/hidden-kosmos>) liegt der Fokus auf der überlieferungsadäquaten Annotation und Edition handschriftlicher Dokumente. Im Projekt „Hidden Kosmos“ werden sämtliche derzeit bekannte Nachschriften zu Alexander von Humboldts weltberühmten Kosmos-Vorträgen (1827/28) erschlossen. Bis dato wurden bereits fünf vollständige, handschriftlich verfasste Nachschriften (Umfang: 2 100 Seiten) im DTA publiziert (siehe www.deutschestextarchiv.de/search/metadata?corpus=avhkv). Bis auf zwei Texte, die zuvor in gedruckten Editionen erschienen waren, kann der größte Teil dieser Nachschriften als der Forschung und der Allgemeinheit unbekannt gelten.

Nun können diese unikalenen Manuskripte ortsunabhängig im DTA studiert, kommentiert, annotiert und heruntergeladen sowie mit Hilfe verschiedener DH-Werkzeuge analysiert werden. Im Vortrag wird unter anderem auf Gemeinsamkeiten und Unterschiede digitaler vs. traditioneller Editorik eingegangen. Der Vergleich der im DTA publizierten, tief strukturierten und annotierten Online-Volltexte mit den früheren Printeditionen macht die Vorzüge digitaler Editionen sichtbar.