

# Werke Chamissos, Humboldts, R. & G. Forsters – und vieler weiterer im Deutschen Textarchiv

## Hochwertige Volltexte, computerlinguistische Analysen und Tools als Grundlage kooperativer Forschung

Im Vortrag werden Primärquellen, die über das DFG-geförderte Projekt „Deutsches Textarchiv“ (DTA, [www.deutschestextarchiv.de](http://www.deutschestextarchiv.de)) der Berlin-Brandenburgischen Akademie der Wissenschaften verfügbar gemacht wurden, und die vielfältigen Möglichkeiten zu deren Nutzung präsentiert. Die Texte des disziplinenübergreifenden, verschiedene Textsorten umfassenden DTA-Kernkorpus wurden nach der frühesten verfügbaren Ausgabe manuell erfasst. Als Ergänzungen wurden hochwertige Volltexte aus externen Quellen kuratiert, in das TEI/XML-basierte DTA-Basisformat<sup>1</sup> konvertiert und in das Gesamtkorpus integriert.<sup>2</sup> Alle Quellen stehen frei über das Internet in verschiedenen Formaten zum Herunterladen und auch zum maschinellen ‚Harvesten‘ über eine API bereit.<sup>3</sup>



Titelblätter einiger im DTA enthaltener Werke Georg Forsters, Reinhold Forsters und A. v. Humboldts

Die akkurate Wiedergabe der historischen Quellen in einem einheitlichen, standardisierten Format, die hohe Textqualität<sup>4</sup> sowie die Zuverlässigkeit der die Objekte beschreibenden Metadaten zeichnen das DTA gegenüber anderen Volltextsammlungen aus. 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. Die automatisierte Erschließung der Volltexte beinhaltet neben der Wortartenbestimmung<sup>5</sup> und der Lemmatisierung der Wortformen<sup>6</sup> auch deren orthographische Normierung. Dabei werden historische Schreibweisen auf ihr modernes Äquivalent abgebildet, was schreibweisentolerante Suchanfragen über das gesamte Korpus ermöglicht.



mehrbändigen Monographien im DTA-Kern- bzw. -Erweiterungskorpus, einer Auswahl von derzeit 164 unselbstständigen Schriften, 21 ‚Abhandlungen‘ nach Akademie-Vorträgen Humboldts sowie einem wachsenden Bestand von Hörernachschriften seiner berühmten ‚Kosmos-Vorträge‘ zusammen.

Innerhalb dieses Korpus lässt sich etwa die Beschäftigung Humboldts mit dem Phänomen des (Erd-)Magnetismus nachverfolgen – von den frühesten Veröffentlichungen der 1790er Jahre<sup>9</sup> über die noch zur Zeit seiner amerikanischen Reise veröffentlichten Beobachtungen in Südamerika<sup>10</sup>, seine 1806 in Berlin angestellten Messungen über den Einfluss des Nordlichts auf die Magnetnadel<sup>11</sup>, die Darstellung des damaligen Wissensstandes im Bereich des Magnetismus in den Kosmos-Vorträgen 1827/28 in der Berliner Universität<sup>12</sup> und vor einem gemischten Publikum in der Singakademie<sup>13</sup>, der Akademievorlesung „Ueber die Mittel, die Ergründung einiger Phänomene des tellurischen Magnetismus zu erleichtern“ aus dem darauffolgenden Jahr<sup>14</sup>, bis hin zur Behandlung dieses Themenfeldes im fünfbändigen *Kosmos*<sup>15</sup>.

Am Beispiel der im DTA verfügbaren Humboldt-Texte wird auch deutlich, dass es nur gemeinsam gelingen kann, die aus Sicht der Forschung in vielen Bereichen und für viele Autoren bzw. Fragestellungen noch zu geringe Zahl der frei im Netz verfügbaren, qualitativ hochwertigen Primärquellen zu erhöhen. Der notwendig kooperative Korpusauf- und -ausbau, die kollaborative Qualitätssicherung und gemeinsame weitere Annotation von Forschungsdaten wird im DTA durch die enge Zusammenarbeit mit mehreren Bibliotheken<sup>16</sup>, Dienstleistern<sup>17</sup>, Universitäten und Forschungsvorhaben<sup>18</sup> umgesetzt.

## Anmerkungen und Referenzen

- <sup>1</sup> Deutsches Textarchiv – DTA-Basisformat – DTABf, [www.deutschestextarchiv.de/doku/basisformat](http://www.deutschestextarchiv.de/doku/basisformat) [Alle URLs in diesem Dokument abgerufen am 1.7.2015], siehe auch Susanne Haaf, Alexander Geyken, Frank Wiegand: The DTA "Base Format": A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources, In: Journal of the Text Encoding Initiative [Online], Issue 8 – PREVIEW | 2014–2015 [<http://jtei.revues.org/1114>; DOI: 10.4000/jtei.1114].
- <sup>2</sup> Umfang des Gesamtkorpus, Stand 18.0.2016: 2825 Dokumente mit ca. 650 000 digitalisierten Seiten, ca. 220 Mio. Tokens, ca. 1,1 Mrd. Zeichen.
- <sup>3</sup> [www.deutschestextarchiv.de/api](http://www.deutschestextarchiv.de/api). Die Volltexte sind jeweils in XML, HTML, plain text und TCF verfügbar.
- <sup>4</sup> Im gesamten Gesamtkorpus bietet das DTA eine Zeichengenauigkeit von mehr als 99,95%. Vgl. dazu Susanne Haaf, Frank Wiegand, Alexander Geyken: Measuring the Correctness of Double-Keying: Error Classification and Quality Control in a Large Corpus of TEI-Annotated Historical Text. In: Journal of the Text Encoding Initiative (JTEI) 4, 2013 [<http://jtei.revues.org/739>; DOI: 10.4000/jtei.739].
- <sup>5</sup> Part of Speech-Tagging (POS), vgl. dazu den Abschnitt Linguistische Analyse historischer Texte (CAB) in der Dokumentation zur Software im Deutschen Textarchiv, <http://www.deutschestextarchiv.de/doku/software#cab>.
- <sup>6</sup> D. h. jedem (normierten) Wort wird eine Grundform (Lemma) zugeordnet; damit wird gewährleistet, dass eine Suche nach dem Term „beilegen“ auch Treffer für die Formen „beygelegt“, „beyzulegen“, „beilegte“ usw. liefert.
- <sup>7</sup> Siehe zu weiteren Möglichkeiten die Hilfe zur Suche in den Korpora des Deutschen Textarchivs, [http://www.deutschestextarchiv.de/doku/DDC-suche\\_hilfe](http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe).

- 
- <sup>8</sup> Vgl. die Übersicht der Humboldt-Texte im DTA unter <http://www.deutschestextarchiv.de/dtae/avh>.
- <sup>9</sup> Vgl. z. B. Humboldt, Alexander von: Neue Entdeckungen [betr. Magnetberg am Fichtelgebirge.] In: Allgemeine Literatur-Zeitung. Intelligenzblatt. Bd. 2 (1797) Nr. 68, Sp. 564-568, [www.deutschestextarchiv.de/humboldt\\_neue\\_1797](http://www.deutschestextarchiv.de/humboldt_neue_1797); Neue Entdeckungen. In: Allgemeine Literatur-Zeitung. Intelligenzblatt. Bd. 3 (1797) Nr. 87, Sp. 722, [www.deutschestextarchiv.de/humboldt\\_entdeckungen\\_1797](http://www.deutschestextarchiv.de/humboldt_entdeckungen_1797); Erklärung. In: Allgemeine Literatur-Zeitung. Intelligenzblatt. Nr. 38 (1797) Sp. 323-326, [www.deutschestextarchiv.de/humboldt\\_erklaerung\\_1797](http://www.deutschestextarchiv.de/humboldt_erklaerung_1797).
- <sup>10</sup> Vgl. z. B. Skizze einer Geologischen Schilderung des südlichen Amerika. In: Allgemeine Geographische Ephemeriden. Bd. 9 (1802) St. 4, S. 310-329, [www.deutschestextarchiv.de/humboldt\\_skizze01\\_1802](http://www.deutschestextarchiv.de/humboldt_skizze01_1802); Notizen Alex. von Humboldt's von seinen Reisen in der Kordillere der Anden und von seinen physikalischen Beobachtungen in Quito und Mexico. In: Annalen der Physik. Bd. 16 (1804) S. 450-493, [www.deutschestextarchiv.de/humboldt\\_notizen\\_1804](http://www.deutschestextarchiv.de/humboldt_notizen_1804).
- <sup>11</sup> [Gilbert, Ludwig Wilhelm]: Die vollständigste aller bisherigen Beobachtungen über den Einfluss des Nordlichts auf die Magnetnadel; angestellt von Herrn Alexander von Humboldt zu Berlin am 20sten Dec. 1806. In: Annalen der Physik. Bd. 29 (1808) S. 425-429, [www.deutschestextarchiv.de/humboldt\\_nordlicht\\_1808](http://www.deutschestextarchiv.de/humboldt_nordlicht_1808).
- <sup>12</sup> Parthey, Gustav: Alexander von Humboldt[.] Vorlesungen über physikalische Geographie. Novmbr. 1827 bis April, 1828. Nachgeschrieben von G. Parthey. [Berlin], [1827/28], [www.deutschestextarchiv.de/parthey\\_msgermqu1711\\_1828](http://www.deutschestextarchiv.de/parthey_msgermqu1711_1828); [N. N.]: Alexander von Humboldts Vorlesungen über physisalische Geographie nebst Prolegomenen über die Stellung der Gestirne. Berlin im Winter von 1827 bis 1828. [Berlin], [1827/28], [www.deutschestextarchiv.de/nn\\_msgermqu2345\\_1827](http://www.deutschestextarchiv.de/nn_msgermqu2345_1827); [N. N.]: Die physikalische Geographie von Herrn Alexander v. Humboldt, vorgetragen im Semestre 1827/28. [Berlin], [1827/28], [www.deutschestextarchiv.de/nn\\_oktavgfeo79\\_1828](http://www.deutschestextarchiv.de/nn_oktavgfeo79_1828).
- <sup>13</sup> Vorlesungen über physicalische Geographie von A. v. Humboldt. [G]eschrieben im Sommer 1829 durch Otto Hufeland. [Berlin], [ca. 1829], [www.deutschestextarchiv.de/hufeland\\_privatbesitz\\_1829](http://www.deutschestextarchiv.de/hufeland_privatbesitz_1829); [N. N.]: Physikalische Geographie. Vorgetragen von Alexander von Humboldt. [Berlin], [1828], [www.deutschestextarchiv.de/nn\\_msgermqu2124\\_1827](http://www.deutschestextarchiv.de/nn_msgermqu2124_1827).
- <sup>14</sup> Humboldt, Alexander von: Ueber die Mittel, die Ergründung einiger Phänomene des tellurischen Magnetismus zu erleichtern. (Auszug aus einer am 2. April 1829 vor der Königlichten Academie der Wissenschaften zu Berlin gehaltenen Vorlesung.) In: Annalen der Physik und Chemie, Bd. 15, St. 3, (1829), S. 319–336, [http://www.deutschestextarchiv.de/humboldt\\_mittel\\_1829](http://www.deutschestextarchiv.de/humboldt_mittel_1829).
- <sup>15</sup> Humboldt, Alexander von: Kosmos. Entwurf einer physischen Weltbeschreibung. Stuttgart u. a., 5 Bde., 1845–62, siehe [www.deutschestextarchiv.de/search?q=Kosmos&in=metadata](http://www.deutschestextarchiv.de/search?q=Kosmos&in=metadata).
- <sup>16</sup> Die Zusammenarbeit mit den Bibliotheken findet v. a. im Bereich Bilddigitalisierung und ggf. auch dem der Metadatenerfassung statt. Vgl. <http://www.deutschestextarchiv.de/doku/kooperationen>.
- <sup>17</sup> In den Bereichen Bilddigitalisierung und Volltexterfassung.
- <sup>18</sup> V. a. in den Bereichen kooperative Edition und tiefere Annotation von Volltexten. Zu nennen sind für die Zusammenstellung des Humboldt-Korpus vor allem die (inzwischen abgeschlossene) Alexander-von-Humboldt-Forschungsstelle der BBAW, in deren Nachfolge das kürzlich begonnene Akademienvorhaben ‚Alexander von Humboldt auf Reisen‘ (<http://www.bbaw.de/forschung/avh-r>), die Humboldt-Universität zu Berlin (Projekt Hidden Kosmos – Reconstructing A. v. Humboldt’s ‚Kosmos-Lectures‘, [www.culture.hu-berlin.de/hidden-kosmos](http://www.culture.hu-berlin.de/hidden-kosmos)) sowie die Universität Potsdam (Lehrstuhl Prof. Dr. Ottmar Ette).