

PROJEKT

Das *Deutsche Textarchiv* (DTA) an der **Berlin-Brandenburgischen Akademie der Wissenschaften** stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit und schafft damit die Grundlage für ein **Referenzkorpus der neuhochdeutschen Sprache**.

ANNOTATION

Der **strukturellen Auszeichnung** aller DTA-Texte liegt das *DTA-Basisformat* (DTABf) zugrunde (Teilmenge von XML TEI/P5). Mit der einheitlichen Textauszeichnung gemäß DTABf wird die Interoperabilität sämtlicher Daten des DTA gewährleistet.

Die **linguistische Annotation** erfolgt automatisch, inklusive Tokenisierung (moot/WASTE), Lemmatisierung, Normalisierung der Schreibungen (DTA::CAB), Part-of-Speech-Tagging.

KORPUSRECHERCHE

Alle Annotationen, Metadaten und Teilkorpora sind über eine **Suchmaschine** (DDC) volltextindiziert und abfragbar. Diese bietet umfangreiche **Trefferansichten, Statistiken, Wortverlaufskurven, diachrone Kollokationen** (DiaCollo) und **vielfältige Exportformate** (JSON, TEI, Ling-XML, TCF, CSV) zur Weiterverarbeitung.

DTA IN CLARIN-D

Textbestand und Metadaten des DTA sind über die **verteilte Infrastruktur von CLARIN-D** zugänglich und können vielfältig nachgenutzt werden. Auf diese Weise wurde eine kollaborative Kultur beim Aufbau und Pflege von Ressourcen innerhalb der Fachgemeinschaften etabliert. Die CLARIN-D-kompatible Infrastruktur des DTA dient darüber hinaus als **dauerhaftes und zentrales Repository für historische Textquellen**.

TEAM

Matthias Boenig, PD Dr. Alexander Geyken, Susanne Haaf, Marius Hug, Dr. Bryan Jurish, Markus Raulf, Christian Thomas, Frank Wiegand
www.deutschestextarchiv.de
redaktion@deutschestextarchiv.de

KORPORA IM DTA

- **DTA-Kernkorpus:** Goldstandard, 1600–1900, Erstausgaben – Belletristik, Wissenschaft, Gebrauchsliteratur: 1 469 Dokumente, 150 Mio Tokens
- **DTA-Erweiterungen (DTAE) :-** 50+ Kooperationen – +Zeitungen, +Handschriften: 4 490 Dokumente, 165 Mio Tokens

Themen	Frauenbewegung	Kolonialismus	
	Texte der ersten Frauenbewegung (TdeF)	Digitale Sammlung dt. Kolonialismus (DSDK)	
Textsorten	Zeitschriften	Zeitungen	Briefe
	Jahrb. des SAC Grenzboten	Neue Rheinische Zt. Hamburg. Corresp.	Sanders Jean Paul
	Poetiken	Funeralschriften	Novellen
	ePoetics	AEDit Oberhofprediger	Novellenschatz
Autoren	Blumenbach	A. v. Humboldt	Gutzkow
	Gesammelte Werke	Vorlesungsnachschriften unselbständige Schriften Briefe	Schriften

FORSCHUNGSPLATTFORM DTAQ

- **Kollaborative Qualitätssicherung:** Lese-/Korrekturtools, Online-Editoren, Ticketsystem, Statistiken, Analysen, frei zugänglich, breite Nutzerbasis
- **Weiterentwicklung** innerhalb von CLARIN-D als Open-Source-Tool

