



# Viele Wege führen ins Basisformat des DTA – Prototypische Kuration historischer Textkorpora zur Integration in Text+

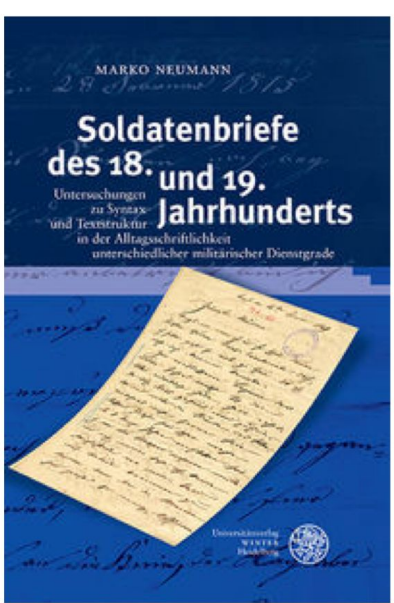
Collections  
Lexical  
Resources  
Editions  
Infrastructure/  
Operations

Marius Hug (Berlin-Brandenburgische Akademie der Wissenschaften)  
Frank Wiegand (Berlin-Brandenburgische Akademie der Wissenschaften)

## Das Deutsche Textarchiv (DTA)

Das DTA ist ein aktives Archiv für deutschsprachige, historische Textsammlungen an der Berlin-Brandenburgischen Akademie der Wissenschaften. Es umfasst annotierte Volltexttranskriptionen von Drucken, Zeitungen und Zeitschriften sowie handgeschriebene Dokumente verschiedener Gattungen und Textarten. Für die Transkriptionen wird mit dem DTABf – ein Subset der TEI-P5-Guidelines – ein etabliertes Basisformat angeboten. Aktuell stellt das DTA etwa 40 Textsammlungen als Forschungsdaten zur Nachnutzung bereit. Im Zentrum steht das DTA-Kernkorpus, das mit rund 1.500 Werken die Grundlage für ein Referenzkorpus des Neuhochdeutschen vom 16. bis frühen 20. Jahrhundert darstellt.

### Office Open XML → DTABf



**Soldatenbriefe des 18. und 19. Jahrhunderts**  
Untersuchungen zu Syntax und Textstruktur in der Alltagsschriftlichkeit unterschiedlicher militärischer Dienstgrade  
Marko Neumann

1. Auflage, 2019  
365 Seiten

ISBN: 978-3-8253-4642-3  
Sortiment: Buch  
Fachgebiet: Germanistik  
Reihe: Germanistische Bibliothek, Band: 66  
Lieferbar: 27.09.2019

Schlagwörter: 18. Jahrhundert, 19. Jahrhundert, Syntax, Textlinguistik, Sozialsemantik, Neuhochdeutsch, Privatbriefe, Briefsteller, Soldatenbriefe, Alltagsschriftlichkeit, Krieg, Briefkorpus, Musterbriefe, Sozialwissenschaft, Militär

52,00 € \*


In den Warenkorb

Rezensionen Dokumente

Anhang zu Marko Neumann: Soldatenbriefe des 18. und 19. Jahrhunderts (PDF 4152kB)

Hintergrund	Dissertation
Datenbasis	Briefkorpus
Umfang	170 Briefe
Format	Office Open XML (.docx)
Sprache	deutsch
Zeitraum	1745–1872
GitHub	deutschestextarchiv/soldatenbriefe
Tools	Perl, XSLT

### ePUB → DTABf



**Die Fortuna von Praeneste**  
Eine Novelle aus dem Jahr 1926

Luise Westkirch (\* 8. Juli 1853 in Amsterdam; † 11. Juli 1941 in München) war eine deutsche Schriftstellerin.

Im antiken Praeneste, dem heutigen Palestrina in Mittelitalien, befindet sich das Heiligtum und Orakel der Fortuna Primigenia. Die Bekanntheit dieser Tempelanlage reicht weit bis in das römische Reich zurück. Laus.


Obwohl diese Region sehr katholisch geprägt ist, bitten die Bürger von Palestrina mit ihren Nöten und Sorgen die Glücksgöttin Fortuna um Hilfe. Dabei leben die Priester der Fortuna aus der vergangenen Zeit scheinbar in den Katzen weiter, die heute im Tempel zu Hause sind. Diese Katzen beeinflussen die Menschen von Palestrina in ihrem Verhalten.

Antonio und Marietta lieben sich. Doch die Beziehung wird durch Luigi gespalten. Über eine Dummheit verliert Antonio seine Marietta, aber Luigi kann die Situation nicht für sich nutzen. Fortuna mit ihren Katzen greift in das menschliche Schicksal ein, der niederträchtige Luigi kommt zu Tode und Antonio und Marietta finden wieder zueinander. Quintessenz: Man kann das Glück nicht zwingen.

Gemeinfreies Werk  
Wiedergefundene Perlen der Literatur Nr. 1, 1. Auflage, 2020

Hintergrund	Public-Private-Partnership
Datenbasis	Textsammlung
Umfang	> 25 Novellen
Format	ePUB
Sprache	deutsch
Zeitraum	1750–1920
GitHub	deutschestextarchiv/stimm-los
Tools	Perl, Python, XSLT

### Reintext → DTABf



Search Upload Communities Log in Sign up

July 4, 2022

**Neuer Deutscher Novellenschatz (1884-1887)**

Weitin, Thomas; Herget, Katharina

Data curator(s)  
Junghenz, Katja; Lemmerich, Julian; Brothager, Judith; Mischner, Vanessa; Schriebe, Ina; Grosser, Laura

Project leader(s)  
Brandes, Ulrik; Weitin, Thomas

Project member(s)  
Olsson, Anastasia; Herget, Katharina; Pipcke, Simon

(English version below)

Der „Neue Deutsche Novellenschatz“ (24 Bände, 1884-1887) wurde von der Digital Humanities Cooperation unter Leitung von Prof. Dr. Thomas Weitin in einem Corrected-OCR-Verfahren im Rahmen des von der VolkswagenStiftung geförderten Projekts *Reading at Scale: Mixing Methods in Literary Analysis* digitalisiert.

420 views 56 downloads See more details...

Indexed in  
**OpenAIRE**

Publication date:  
July 4, 2022  
DOI:  
10.5281/zenodo.619371  
Keyword(s):

Hintergrund	Drittmittelprojekt
Datenbasis	Zenodo-Korpus
Umfang	70 Novellen
Format	Reintext
Sprache	deutsch
Zeitraum	1884–1887
GitHub	deutschestextarchiv/nschatz_deu
Tools	Perl, Python, XSLT

Zertifizierte, nachnutzbare, etablierte Infrastruktur und Dokumentation: DTA – DWDS – Text+

Soldatenbriefe des 18. und 19. Jahrhunderts

1745–1872

Link Landing-Page Link Suche

Sprache deu Format DTABf Format TCF

Faksimiles nein Transkription manually

Dokumente 170 Sätze 3319 Tokens 96023

Genre Gebrauchsliteratur::Brief Lizenz CC BY-SA 4.0

Das Korpus „Soldatenbriefe“ umfasst 170 Briefe von Offizieren, Unteroffizieren und einfachen Soldaten, adressiert an die Familien in der Heimat. Die Briefe wurden im Zeitraum von 1745 bis 1872 auf Deutsch verfasst; ein deutlicher Schwerpunkt liegt auf Briefen aus den Koalitions- und Befreiungskriegen (1792–1815), dem Deutschen Krieg (1866) und dem Deutsch-Französischen Krieg (1870/71).

Datensatz

stimm-los. Wiedergefundene Perlen der Literatur

1750–1920

Link Landing-Page Link Suche

Sprache deu Format DTABf Format TCF

Faksimiles nein Transkription manually

Dokumente 28 Sätze 20662 Tokens 488727

Genre Belletristik::Novelle Lizenz CC BY-SA 4.0

Das Korpus „stimm-los“ enthält vornehmlich Novellen aus dem Zeitraum 1750–1920. Durch die Transkription dieser Texte verfolgt stimm-los das Ziel, in Fraktur gesetzte Klassiker der Weltliteratur verschiedener Sprachen einer breiten Leserschaft zugänglich zu machen. Das DTA übernimmt die Konvertierung der Werke in das Format DTABf sowie die Integration und langfristige Bereitstellung der Resource in der DTA-Infrastruktur.

Datensatz

Neuer Deutscher Novellenschatz (1884–1887)

1884–1887

Link Landing-Page Link Suche

Sprache deu Format DTABf Format TCF

Faksimiles ja Transkription corrected ocr

Dokumente 70 Sätze 81848 Tokens 1,7 Mio

Genre Belletristik::Novelle Lizenz CC BY-SA 4.0

Das Korpus „Neuer Deutscher Novellenschatz“ enthält eine von Paul Heyse und Ludwig Laistner herausgegebene Novellensammlung mit 70 Texten von 66 Autorinnen bzw. Autoren. Die Sammlung hat einen zeitgenössisch-realistischen Schwerpunkt. Mit wenigen Ausnahmen stammen die Texte aus der zweiten Hälfte des 19. Jahrhunderts, ein Großteil der aufgenommenen Novellen ist heute dekanonisiert.

Datensatz

Einstieg deutschestextarchiv.de/sammlungen/soldatenbriefe  
Suche www.dwds.de/d/korpora/soldatenbriefe  
Blog textplus.hypotheses.org/4815

Einstieg deutschestextarchiv.de/sammlungen/stimm-los  
Suche www.dwds.de/d/korpora/stimm-los  
Blog textplus.hypotheses.org/6193

Einstieg deutschestextarchiv.de/sammlungen/nschatz\_deu  
Suche www.dwds.de/d/korpora/nschatz\_deu  
Blog textplus.hypotheses.org/7549