

Ressourcenbeitrag: Deutsches Textarchiv (DTA)

Jan Horstmann ¹

Alexandra Kern

1. Universität Münster

forTEXT

Thema:	Bibliografie	DOI:	10.48694/fortext.3791
Jahrgang:	1	Ausgabe:	11
Erscheinungsdatum:	30-11-2024	Erstveröffentlichung:	2018-07-24 auf forttext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte **Begriffe** werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

1. Kurzbeschreibung

Das Deutsche Textarchiv (DTA) ist ein digitales Volltextkorpus (vgl. **Korpus**) historischer Druckwerke zwischen 1600 und 1900 und eignet sich als eine hochwertige Quelle für zitierfähige Primärtexte.



Abb. 1: Benutzeroberfläche des DTA

Steckbrief

- <http://www.deutschestextarchiv.de>
- Volltextsammlung: 4422 deutschsprachige Druckwerke zwischen ca. 1600 und ca. 1900 (aktuelle Zahlen unter: <http://www.deutschestextarchiv.de/doku/ueberblick#umfang4>)
- Ausgaben: vorzugsweise Erstveröffentlichungen, ggf. historisch-kritische
- Textsorten: Zeitung, Gebrauchsliteratur, Wissenschaft, Belletristik; gedruckte und handschriftliche Vorlagen
- **Metadaten:** Titel, Autor, Herausgeber, Übersetzer, Ort, Verlag, Auflage, Band (DTA-Basisformat)
- Projekt der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), gefördert von der Deutschen Forschungsgemeinschaft (DFG) von 2007 bis 2016; Erweiterung im Rahmen von CLARIN-D von 2017 bis 2020
- Ziel: Abbildung der sprachhistorischen Entwicklungen der deutschen Sprache seit dem Ende der frühneuhochdeutschen Sprachperiode
- Downloadformate: **XML (TEI P5)**, **HTML**, Text, TCF (text annotation layer), TCF (tokenisiert (vgl. **Type/Token**)), serialisiert, lemmatisiert (vgl. **Lemmatisieren**), normalisiert; weitere Downloadformate für Metadaten sowie spezifische Downloadformate für Tools (wie etwa Voyant (Flüh 2024))

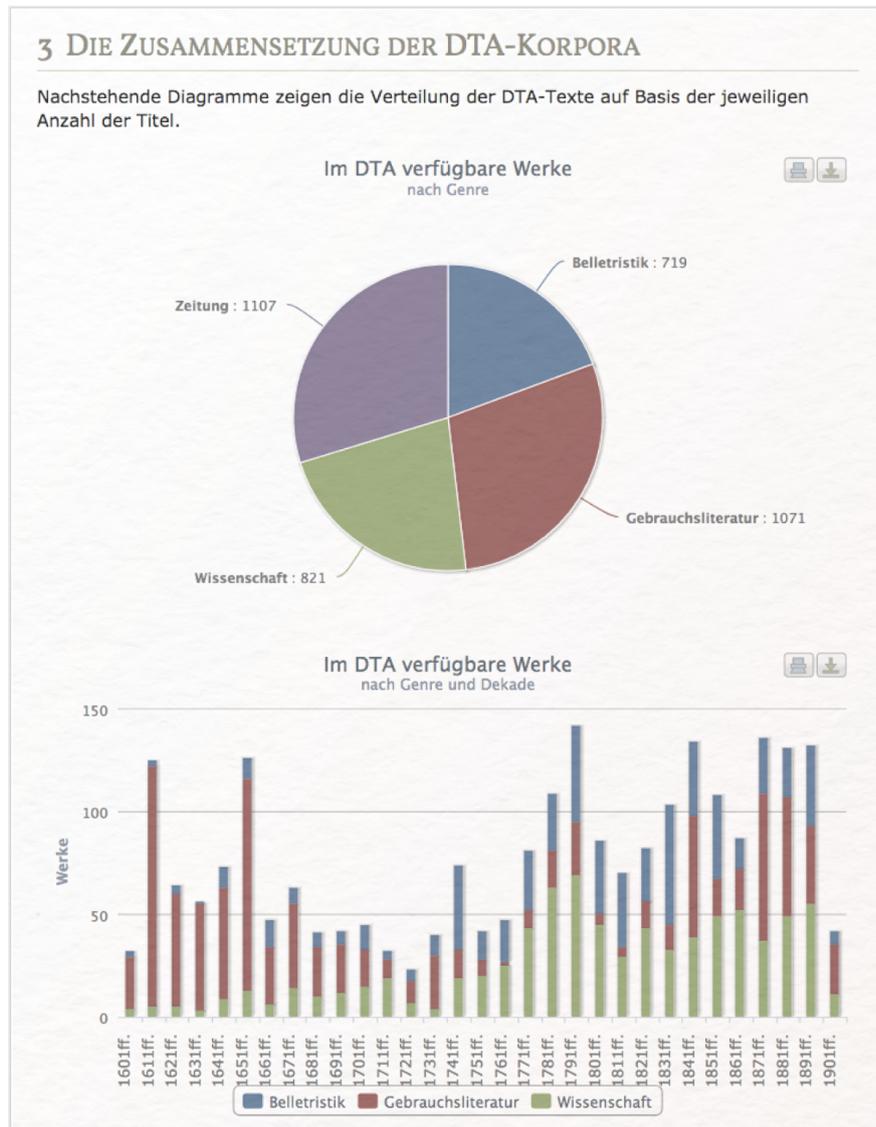


Abb. 2: Werke im DTA

2. Anwendungsbeispiel

Sie wollen digital unterstützt das Frauenbild in Goethes *Iphigenie auf Tauris* analysieren. Eine wesentliche Voraussetzung für diesen Anwendungsfall ist, dass Ihnen ein (idealerweise zitierfähiger) digitalisierter Primärtext vorliegt. Je korrekter die zugrundeliegende Textquelle, desto genauer sind auch die an ihr durchgeführten digitalen Methoden der Textanalyse und **Annotation**. Für das Auffinden eines solchen Textes bietet sich die Suche im Deutschen Textarchiv (DTA) an. Im DTA sind Erstveröffentlichungen und historisch-kritische Ausgaben von Druckwerken, Zeitungen und Manuskripten insbesondere aus der Zeit von 1600 bis 1900 vorhanden, die über eine Schnellsuchfunktion (vgl. **Query**) unproblematisch gefunden und **open access** (vgl. **Open Access**) in unterschiedlichen Dateiformaten heruntergeladen werden können. Die Textsammlung wurde sorgfältig zusammengestellt und beinhaltet zahlreiche Werke einer Vielzahl von Autor*innen.

3. Diskussion

3.1 Kann ich das DTA für wissenschaftliche Arbeiten nutzen?

Ja – denn die hohe Textqualität der DTA-Texte wird u. a. dadurch erreicht, dass sowohl formativ (d. h. vor der Texterfassung) als auch summativ (d. h. nach der Texterfassung) Qualitätskontrollen durchgeführt werden. Das DTA bemüht sich außerdem darum, den historischen Sprachstand der Werke zu bewahren. Deshalb werden möglichst Erst- bzw. frühe Original-Ausgaben der Texte zugrunde gelegt, die zudem strukturell in Kapitel, Unterkapitel und Absätze unterteilt werden. Dargestellt wird jeweils das originale Dokument als Bild in hoher

Auflösung und eine entsprechende elektronische Version des Textes.
 Die Überführung der für das DTA ausgewählten Texte in elektronische Form erfolgte für 200 Texte im automatischen Verfahren **OCR** und für ca. 1300 Texte im manuellen Verfahren, bei dem der Text zunächst von Nicht-Muttersprachlern eingegeben wird, um anschließend auf eventuelle Abweichungen hin überprüft zu werden (Double-Keying-Verfahren (vgl. **Double-Keying**); s. Möglichkeiten der Textdigitalisierung (Horstmann 2024)). Auch die 200 zunächst automatisch erfassten Texte wurden anschließend manuell korrigiert, um sie dem Qualitätsstandard des DTA anzugleichen. Die Texterfassung und -aufbereitung wird somit von einer ständigen Überprüfung begleitet, die nur dann Korrekturen vornimmt, wenn fehlerhafte Eingaben erkannt wurden oder das historische Dokument lückenhaft bzw. verfälscht ist.
 Die formative Qualitätskontrolle umfasst schließlich auch eine ständige Pflege und Überarbeitung der generellen DTA-Richtlinien zur Texterfassung und -annotation, die im Menüpunkt „Dokumentation“ verfolgt werden können.
 Die Überprüfung der digitalisierten Dokumente endet jedoch nicht mit den hier beschriebenen formativen Qualitätskontrollen, sondern wird summativ durch die webbasierte, kollaborative DTA-Qualitätssicherung (DTAQ) fortgesetzt (siehe Abb. 3). DTAQ ermöglicht registrierten Benutzer*innen, eigene Annotationen und Korrekturen im jeweiligen Text anzumerken, die vom DTA-Team kontrolliert und gegebenenfalls eingearbeitet werden. Außerdem können auch neue Texte eingespeist werden, die wiederum den aufgestellten Richtlinien des DTA für die Texterfassung und Annotation entsprechen müssen. Für ein solches Unterfangen stehen auf der Webseite Vorlagen bereit, die Ihnen anschaulich die einzelnen Schritte aufzeigen.

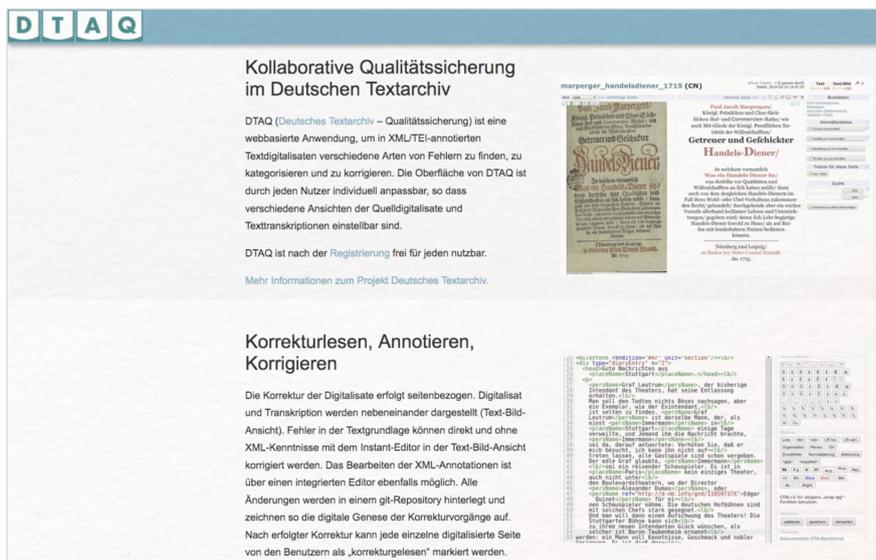


Abb. 3: Kollaborative Qualitätssicherung im DTA: DTAQ

3.2 Wie benutzerfreundlich ist die Arbeit mit dem DTA?

Das DTA bietet Ihnen eine klar strukturierte, aber komplexe Arbeitsumgebung. In ihren Grundzügen kann sie jedoch auch von Erstnutzer*innen ohne größere Vorkenntnisse durch Ausprobieren erschlossen werden. Auf der Homepage finden Sie zunächst eine horizontale Navigation mit den Menüpunkten „Texte“, „Projekt“, „Dokumentation“ und „Impressum“. Die Menüpunkte gliedern sich wiederum in mehrere Unterpunkte, wie beispielsweise „DTA-Leitlinien“, „DTA-Textauswahl“ und „DTA-Quellen“. Einerseits bringt die große Menge an Unterpunkten den Vorteil mit sich, die Dokumentation der Textauswahl und -aufbereitung ausführlich nachvollziehen zu können, andererseits droht jedoch die Gefahr, dass unerfahrene Nutzer*innen schnell die Übersicht verlieren und nur schwer die wesentlichen Bedienschritte für eine Erstnutzung finden.
 Unterstützung bei der Bedienung und Nutzung des DTA finden Sie unter dem Menüpunkt „Dokumentation“, Unterkategorie „Hilfe“. Hier werden u. a. die DTA-Leitlinien, die DTA-Richtlinien zur Texterfassung, das DTA-Basisformat, die sog. Korrekturfibel (für eigene Anmerkungen und Korrekturen) sowie die verschiedenen Ansichten der Texte komprimiert erklärt.
 Mögliche Fragen bezüglich der Bedienung und Nutzung der Suchmaschinen bleiben dennoch unbeantwortet. Diese können jedoch in den Unterpunkten „linguistische Suche“ bzw. „Projektüberblick“ oder unter dem Button „Hilfe“ (neben der Navigation) nachgelesen werden. Nachteil an einer solchen Form der Darstellung ist, dass vorrangig linguistische Suchanfragen bzw. die Nutzung der systemimmanenten Suchmaschine DDC beschrieben und mit Hilfe von Beispielen veranschaulicht werden. Unerfahrene Nutzer*innen können schnell durch die ausführlichen Beschreibungen verunsichert werden und erhalten keine konkreten Antworten auf Belange, die nicht linguistischer Natur sind.

Insgesamt erfordert die Bedienung der Schnellsuchfunktion, die vor allem für die Recherche nach bestimmten Texten nützlich ist, jedoch keine größeren Vorkenntnisse und kann auch von Erstnutzer*innen durch eigenes Ausprobieren erschlossen werden. Die Volltexte selbst werden seitenweise sowohl als Bild der originalen Vorlage als auch als HTML-Version dargestellt, die den historischen Text in eine moderne Schriftart überführt hat, sodass auch Texte, die ursprünglich in einer Frakturschrift oder handschriftlich veröffentlicht wurden, für alle Nutzer*innen lesbar sind.

Sollten sich bei der Nutzung größere Schwierigkeiten ergeben, können Sie das Team des DTA entweder schriftlich über ihre Mail-Adresse (Menüpunkt „Impressum“) kontaktieren oder eine der vom DTA angebotenen Schulungen besuchen. Bei den Schulungen erhalten Sie einen Einblick über die DTA-Erfassungsrichtlinien sowie das DTA-Basisformat und lernen über praxisnahe Beispiele den Umgang mit der Suchmaschine DDC. Bei Bedarf bietet das DTA zudem an, Transkriptions- und Annotationsarbeiten kontinuierlich zu begleiten und zu kontrollieren.

4. Wie funktioniert die Textsuche im DTA?

Sind Sie noch unschlüssig, welchen Primärtext Sie für Ihre Untersuchung heranziehen wollen, können Sie ohne viel Aufwand in der Textsammlung des DTA stöbern. Hierfür müssen Sie lediglich den Menüpunkt „Texte“, Unterkategorie „Verfügbar“ oder „Zeitleiste“, auswählen, um eine alphabetisch oder zeitlich sortierte Auflistung (siehe Abb. 4) der Druckwerke zu erhalten.

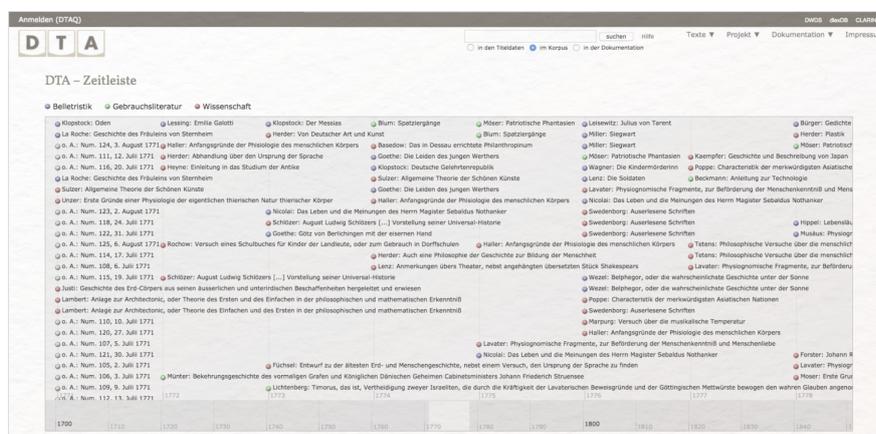


Abb. 4: Zeitleiste der Texte im DTA

Möchten Sie jedoch einen bestimmten Primärtext auffinden, empfiehlt es sich, die Schnellsuchfunktion auf der Startseite links neben der Navigation zu nutzen. In das Suchfeld können Sie den Autor*innennamen oder den Titel des Druckwerkes eingeben. Vor der Suche sollten Sie unter dem Suchfeld festlegen, ob Sie die „Titeldaten“, das „Korpus“ oder die „Dokumentation“ durchsuchen wollen. Suchen Sie nach (literarischen) Primärtexten, sollten Sie entweder die Titeldaten oder das Korpus auswählen. Beide Fundorte leiten Sie zu einer neuen Ansichtssseite weiter, auf der die Suchergebnisse dargestellt werden. Die Informationen zu den einzelnen Suchergebnissen führen die wesentlichen Metadaten wie den Titel, die Auflage, den Autor*innennamen, das Erscheinungsjahr bzw. den -ort und den Umfang des Druckwerkes auf. Über einen Doppelklick auf den Titel kann das Druckwerk auf einer eigenen Ansichtssseite vollständig eingesehen werden.

Sollte bei den ersten Suchergebnissen das von Ihnen gesuchte Werk nicht dabei sein, können Sie Ihre Suche über weitere Angaben verfeinern. Für ein solches Unterfangen stehen Ihnen oben auf der Ansichtssseite der Suchergebnisse Filteroptionen, wie der Autor*innennamen, der Titel, die Klassifikation, der Druckort, der Verlag oder auch das Erscheinungsjahr des Werkes zur Verfügung. Die einzelnen Filteroptionen sind zudem miteinander kombinierbar. Bei einer erfolgreichen Suche kann das Druckwerk in unterschiedlichen Formaten (verschiedene XML-, HTML-, Text- und TCF-Formate) und sogar in normalisierter Orthografie (was für einige Verfahren des Distant Reading (vgl. **Distant Reading**) sinnvoll sein kann) heruntergeladen und weiterverwendet werden.

Sie sind etwas experimentierfreudiger? Dann kann es zudem lohnend sein, sich genauer mit der Funktion der linguistischen Suchmaschine DDC zu beschäftigen. Sie ermöglicht es, nach einer exakten Wortform, einer flektierten (z. B. die Anfrage nach „sprach“ liefert u. a. die Ergebnisse für „sprechen“, „spricht“, „gesprochen“) oder graphematischen Variante für ein Wort (z. B. „Kleid“: „Kleidt“, „Kleydt“, „Cleyd“, „Cleit“ etc.) zu suchen und erleichtert es so, die sprachliche Ausgestaltung eines Textes zu erfassen.

Externe und weiterführende Links

- Deutsches Textarchiv: <https://web.archive.org/web/20241106121355/https://www.deutschestextarchiv.de/> (Letzter Zugriff: 06.11.2024)
- Volltextsammlung Überblick: <https://web.archive.org/web/20241106121508/https://www.deutschestextarchiv.de/doku/ueberblick#umfang> (Letzter Zugriff: 06.11.2024)

Bibliographie

- Flüh, Marie. 2024. Toolbeitrag: Voyant. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3775, <https://fortext.net/tools/tools/voyant>.
- Geyken, Alexander, Matthias Boenig, Susanne Haaf, Bryan Jurish, Christian Thomas und Frank Wiegand. 2018. Das Deutsche Textarchiv als Forschungsplattform für historische Daten in CLARIN. In: *Digitale Infrastrukturen für die germanistische Forschung*, hg. von Henning Lobin, Roman Schneider, und Andreas Witt, 219–248. Berlin, Boston: de Gruyter. doi: 10.1515/9783110538663-011,.
- Geyken, Alexander und Thomas Gloning. 2015. A living text archive of 15th-19th-century German. Corpus strategies, technology, organization. In: *Historical Corpora. Challenges and Perspectives*, hg. von Jost Gippert und Ralf Gehrke, 165–180. Tübingen: Narr.
- Haaf, Susanne. 2017. Das DTA-Basisformat in neuem Gewand. *Im Zentrum Sprache. Untersuchungen zur deutschen Sprache in Geschichte und Gegenwart*. <https://sprache.hypotheses.org/147> (zugegriffen: 13. Juli 2018).
- Haaf, Susanne, Alexander Geyken und Frank Wiegand. 2015. The DTA „Base Format“: A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources. *Journal of the Text Encoding Initiative*, Nr. 8. doi: 10.4000/jtei.1114,.
- Haaf, Susanne und Christian Thomas. 2016a. Enabling the Encoding of Manuscripts within the DTABf: Extension and Modularization of the Format. *Journal of the Text Encoding Initiative (JTEI)*, Nr. 10. doi: 10.4000/jtei.1650, <https://journals.openedition.org/jtei/1650>.
- . 2016b. Die Historischen Korpora des Deutschen Textarchivs als Grundlage für sprachgeschichtliche Forschungen. In: *Sprachgeschichte des Deutschen: Positionierungen in Forschung, Studium, Schule*, hg. von Holger Runow, Volker Harm, und Levke Schwiek, 217–234. Stuttgart: Hirzel.
- Horstmann, Jan. 2024. Methodenbeitrag: Möglichkeiten der Textdigitalisierung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 3. Textdigitalisierung und Edition (12. Juni). doi: 10.48694/fortext.3741, <https://fortext.net/routinen/met-hoden/moeglichkeiten-der-textdigitalisierung>.
- Wiegand, Frank, Christian Thomas, Susanne Haaf, Alexander Geyken, Bryan Jurish und Matthias Boenig. 2018. Recherchieren, Arbeiten und Publizieren im Deutschen Textarchiv: ein Praxisbericht. *Zeitschrift für Germanistische Linguistik* 46, Nr. 1: 147–161. doi: 10.1515/zgl-2018-0009,.

Glossar

Annotation Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.

Browser Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.

Close Reading Close Reading bezeichnet die sorgfältige Lektüre und Interpretation eines einzelnen oder weniger Texte. Close Reading ist in der digitalen Literaturwissenschaft außerdem mit der manuellen **Annotation** textueller Phänomene verbunden (vgl. auch **Distant Reading** als Gegenbegriff).

Distant Reading Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computationale Verfahren auf häufig große Mengen an Textdaten angewandt werden, ohne dass die Texte selber gelesen werden. Meist stehen hier quantitative Analysen im Vordergrund, es lassen sich jedoch auch qualitative **Metadaten** quantitativ vergleichen. Als Gegenbegriff zu **Close Reading** wurde der Begriff insbesondere von Franco Moretti (2000) geprägt.

Double-Keying Double-Keying ist eine Variante des **Keying**, bei der zwei Personen den Inhalt eines Dokumentes abtippen. Anschließend sucht ein Computerprogramm nach Differenzen zwischen den beiden Versionen. Gefundene Tippfehler werden dann von einer dritten Person korrigiert. So entstehen nahezu fehlerfreie Textdigitalisate.

HTML HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben

die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.

Keying In den Bibliotheks- und Textwissenschaften beschreibt Keying das manuelle Erfassen, also das Abtippen, eines Textes im Zuge seiner Digitalisierung (siehe auch **Double-Keying**).

Korpus Ein Textkorpus ist eine Sammlung von Texten. Korpora (Plural für „das Korpus“) sind typischerweise nach Textsorte, Epoche, Sprache oder Autor*in zusammengestellt.

Lemmatisieren Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.

Machine Learning Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwendet werden.

Markup (Textauszeichnung) Die Textauszeichnung (eng. *Markup*) fällt in den Bereich der Daten- bzw. Textverarbeitung, genauer in das Gebiet der Textformatierung, welche durch **Auszeichnungssprachen** wie **XML** implementiert wird. Dabei geht es um die Beschreibung, wie einzelne Elemente eines Textes beispielsweise auf Webseiten grafisch dargestellt werden sollen.

Markup Language Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z. B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.

Metadaten Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.

Named Entities Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.

OCR OCR steht für *Optical Character Recognition* und bezeichnet die automatische Texterkennung von gedruckten Texten, d. h. ein Computer „liest“ ein eingescanntes Dokument, erkennt und erfasst den Text darin und generiert daraufhin eine elektronische Version.

Open Access Open Access bezeichnet den freien Zugang zu wissenschaftlicher Literatur und anderen Materialien im Internet.

POS PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.

Preprocessing Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.

Query *Query* bedeutet „Abfrage“ oder „Frage“ und bezeichnet eine computergestützte Abfrage zur Analyse eines Textes. Um Datenbestände zu durchsuchen, werden Abfragesprachen eingesetzt, die *Queries* (Anfragen) an den Datenbestand senden. So bilden alle möglichen Queries zusammen die *Query Language* eines Tools.

TEI Die *Text Encoding Initiative* (TEI) ist ein Konsortium, das gemeinsam einen Standard für die Darstellung von Texten in digitaler Form entwickelt. Die TEI bietet beispielsweise Standards zur Kodierung von gedruckten Werken und zur Auszeichnung von sprachlichen Informationen in maschinenlesbaren Texten (siehe auch **XML** und **Markup**).

Type/Token Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.

Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.

XML XML steht für *Extensible Markup Language* und ist eine Form von **Markup Language**, die sowohl computer- als auch menschenlesbar und hochgradig anpassbar ist. Dabei werden Textdateien hierarchisch strukturiert dargestellt und Zusatzinformationen i. d. R. in einer anderen Farbe als der eigentliche (schwarz gedruckte) Text dargestellt. Eine standardisierte Form von XML ist das **TEI-XML**.