

Methodenbeitrag: Digitales Bibliografieren

Marie Flüh  ¹

1. Universität Hamburg

forTEXT

Thema:	Bibliografie	DOI:	10.48694/fortext.3786
Jahrgang:	1	Ausgabe:	11
Erscheinungsdatum:	30-11-2024	Erstveröffentlichung:	2019-03-04 auf fortext.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. Definition

Der Begriff Bibliografie wird im Deutschen in dreifachem Sinn gebraucht und bezeichnet das Literaturverzeichnis (weitere synonyme Ausdrücke sind Bibliothek, Lexikon, Index oder Katalog), Bibliografieren als Herstellung des Literaturverzeichnisses sowie die Lehre der Herstellung und Nutzung von Literaturverzeichnissen (Bartsch 1989, 16). Digitales Bibliografieren meint sowohl die digital gestützte Suche nach als auch die Organisation von Forschungsliteratur z. B. mithilfe eines digitalen Literaturverwaltungsprogramms. Solche multifunktionalen Tools begleiten – von der Literaturrecherche in elektronischen Bibliografien über den Schreibprozess unter Einbezug der Sekundärliteratur bis zur Publikation – sämtliche Schritte des literaturwissenschaftlichen Arbeitens in digital unterstützender Funktion.

2. Anwendungsbeispiel

Sie möchten eine kollaborative wissenschaftliche Arbeit über Christoph Martin Wieland und die Entstehung des Bildungsromans anfertigen und sind auf der Suche nach allen für das Thema wesentlichen Literaturdokumenten. Die Ergebnisse möchten Sie mit Ihren Kolleg*innen teilen und ein Literaturverzeichnis nach Harvard-Zitierweise erstellen. Digitale Bibliografiertools können Sie bei diesen Schritten unterstützen.

3. Literaturwissenschaftliche Tradition

Ein Wesensmerkmal der geisteswissenschaftlichen Arbeitspraxis stellt der unablässige Umgang mit Literaturdokumenten (Primär- und Sekundärliteratur) dar (Paschek 1999, 13), die in Bibliografien verzeichnet werden. Die Germanistik als Wissenschaft von der deutschen Sprache und Literatur blickt auf eine vergleichsweise kurze Tradition zurück. Die eigentliche Forschung setzte erst in der Blütezeit der Romantik um 1850 u. a. mit den Arbeiten der Gebrüder Grimm ein. Die daraus resultierenden Publikationen liegen folglich kaum mehr als 160 Jahre zurück und sind in Bibliografien verzeichnet, die noch jüngeren Datums sind (Raabe 1980, 4). Bibliografieren als aktive Praxis und Ermittlung von Literaturdokumenten (Jeßing und Köhnen 2012, 372; Jeßing 2017, 33) blickt folglich genau wie die Fachdisziplin selbst auf eine kurze und erst im ausgehenden 19. Jahrhundert einsetzende Tradition zurück.

Ein Blick auf die Etymologie verdeutlicht, dass das Bibliografieren als Technik des wissenschaftlichen Arbeitens in anderen Kulturen jedoch eine weitaus längere Tradition hat. Die Wortfamilie um den Begriff der Bibliografie wurde zunächst ausschließlich in der antiken und mittelalterlichen Literatur der Griechen für die Tätigkeit des Schreibens oder Abschreibens von Büchern verwendet. Einen bibliografischen Sinn, der unserem heutigen Gebrauch entspricht, besaß das griechische Wort in der Antike und im Mittelalter zunächst nicht. Im Lateinischen fehlte eine Übernahme oder sinngemäße Nachbildung zunächst gänzlich. Erst im 17. Jahrhundert wurde der Term Bibliografie als Bezeichnung für Bücherverzeichnisse verwendet und löste deren Benennung als Index, Elenchus, Bibliotheca oder Cataloga durch den Begriff der Bibliografie ab (Bartsch 1979, 182). Da die Informationsfülle und das Wachstum von Wissenschaften, Technik und Verwaltung maßgebliche Kennzeichen dieser Zeit darstellen, entwickelten sich mannigfaltige Spezial- und Fachbibliografien (vgl. ebd., 199). Die bis heute fortlaufende Entwicklung von Nationalbibliografien setzte in den mitteleuropäischen Ländern gegen Mitte des 19. Jahrhunderts ein (vgl. ebd., 207). Den gesetzlichen Auftrag zur bibliografischen Verzeichnung sämtlicher in Deutschland erscheinenden Veröffentlichungen erfüllt die Deutsche Nationalbibliothek (DNB).

4. Diskussion

Während vormalig Bibliografien ausschließlich in gedruckter Form vorlagen und in einzelnen Schritten händisch unter Berücksichtigung der geltenden Zitierweisen erstellt wurden, gibt es heute zahlreiche digitale Literaturverwaltungsprogramme, die bei der Literaturrecherche sowie bei der digitalen Literatur- und Quellenverwaltung assistieren und das steigende elektronische Informationsangebot der Bibliotheken auffangen. Die analoge Suche nach Sekundärliteratur in gedruckten Bibliografien wird durch die digitale Recherche in den elektronischen Katalogen der Bibliotheken (Online Public Access Catalogues, OPACs (vgl. OPAC)), digitalen Fachportalen und Fachbibliografien ergänzt oder gar ersetzt. Die zu diesem Zweck entwickelten leistungsfähigen und multifunktionalen Tools (wie bspw. *Citavi*, *Mendeley*, *EndNote* oder *Zotero* (Flüh 2024a)) sind dabei mehr als persönliche Literaturdatenbanken. Ihre Funktionen sind vielmehr auf die Gesamtheit der Arbeitsschritte eines geisteswissenschaftlichen Forschungsvorhabens abgestimmt: Recherche von Forschungsliteratur, Erstellen von Bibliografien, Zitation der Sekundärliteratur und schlussendlich die Publikation der Arbeit werden durch solche Tools digital unterstützt.

Für die Beschaffung und Erarbeitung der Primärliteratur, die den Ausgangspunkt des literaturwissenschaftlichen Arbeitens darstellt (Jeßing 2010, 11; Jeßing 2017, 33), und die Suche nach digitalen Textkorpora (vgl. **Korpus**) deutschsprachiger Primärliteratur, bieten sich bspw. das **Deutsche Textarchiv** oder das **TextGrid Repository** an (siehe auch Horstmann und Kern (2024) zum Deutschen Textarchiv (DTA) und Horstmann (2024) zum TextGrid Repository). Beide Langzeitarchive kooperieren mit dem Fachportal Germanistik im Netz (GiN) (Flüh 2024b), in dem sich die Suche nach Primär- und Sekundärliteratur verbinden lässt (TextGrid: bereits Bestandteil; DTA: Einbezug im Rahmen der laufenden Erneuerung von GiN). Um digitalisierte Bestände alter Drucke, Handschriften, Zeitschriften oder Dokumente aus Nachlässen zu erhalten, stellen außerdem die fachspezifischen Sammlungen der Universitätsbibliotheken eine wichtige erste Anlaufstelle dar.

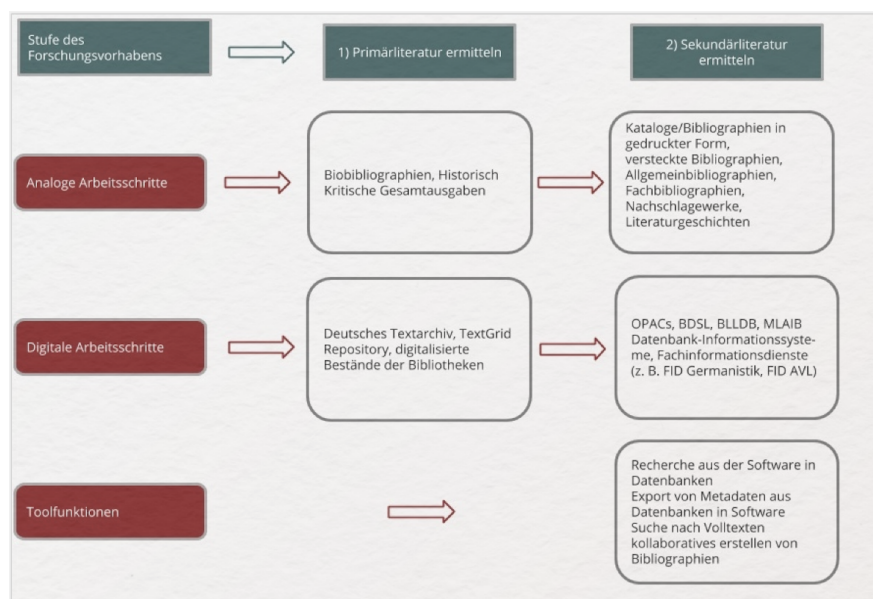


Abb. 1: Systematischer Verlauf eines Forschungsvorhabens: Arbeitsschritte 1 und 2

An zweiter Stelle des literaturwissenschaftlichen Arbeitens (siehe Abbildung 1) steht die extensive Beschaffung von Forschungsliteratur zu einem literarischen Text bzw. Korpus und/oder der entsprechenden Forschungsfrage (Sekundärliteratur: Hintergrundinformationen zu Autor*in und Text, zu seinem/ihrem Gesamtwerk, zur literaturgeschichtlichen Epoche, zu methodologischen Konzepten, zur literarischen Gattung, Stand der Forschung) (Jeßing 2010, 11). Literaturverwaltungsprogramme begleiten Ihre Recherche in unterstützender Funktion, indem sie als Schnittstelle zu einer Vielzahl an Online-Datenbanken fungieren und den Export von Referenzen aus diesen Datenbanken ermöglichen. Die kommerzielle Literaturverwaltung des amerikanischen Informationsdienstleisters Thomson-Reuters *EndNote* (aktuell in der Version X9 für Windows und Mac erhältlich) oder das in der Basisversion kostenlos erhältliche Tool *Citavi* (aktuell in der Version *Citavi Free 6.3* für Windows 7, 8, 10 erhältlich) gestatten aus der Software hinaus die Recherche in OPACs, Fachbibliografien und weiteren vordefinierten Datenbanken (*Citavi 6.1*: über 4500 vordefinierte Zugänge; *EndNote X9*: über 6000 vordefinierte Zugänge, von denen 400 vorinstalliert sind). Bei Bedarf können weitere Datenbanken oder Kataloge bei den Herstellern angefordert und heruntergeladen werden.

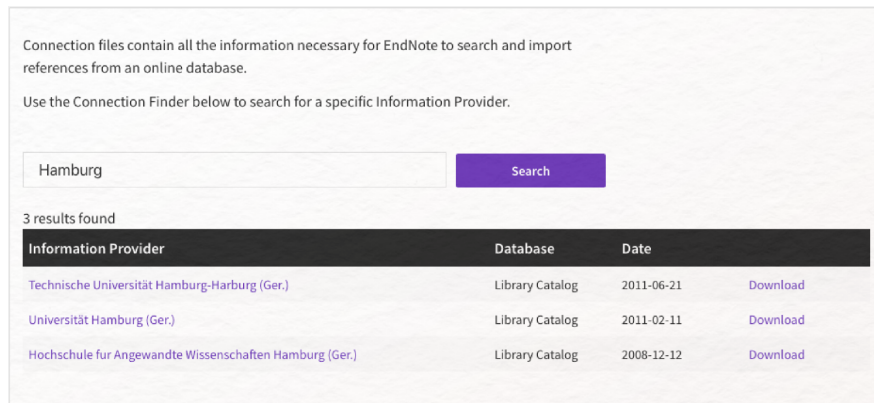


Abb. 2: Download weiterer Datenbanken in das Literaturverwaltungsprogramm EndNote über die Homepage der Betreiber

Im Rahmen der Recherche ermöglichen Tools wie *EndNote* (siehe Abbildung 2) den Import von Dateien über einen Importfilter und den direkten Export von **Metadaten** aus Datenbanken in die Software. Auf diese Weise lassen sich Dateien aus bibliografischen Datenbanken und anderen Literaturverwaltungsprogrammen direkt in die Software einspeisen. Importierte Dateien können den einzelnen Beiträgen als Anhang (z. B. als PDF Dokument (vgl. **PDF**), als Link zur **URI** oder als Link auf eine Datei) hinzugefügt werden.

In vielen Fällen können die gängigen Literaturverwaltungsprogramme direkt auf die Online-Kataloge von Universitätsbibliotheken zugreifen (siehe Abbildung 3), bibliografische Daten aus den OPACs herunterladen und beispielsweise in die eigene *Citavi*-, *EndNote*- oder *Zotero*-Bibliothek importieren.

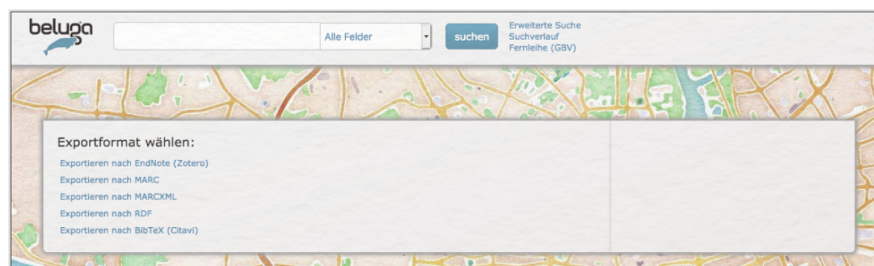


Abb. 3: Der Export von Metadaten über den Katalog der Hamburger Universitäten

Sie können Metadaten außerdem über einen Button in der Browserleiste (vgl. **Browser**) direkt in die Literaturverwaltungsprogramme importieren. Voraussetzung hierfür ist die Installation eines Web-Importers (*Zotero*: *Zotero-Connect*, siehe Abbildung 4), *Mendeley*: *Web Importer*, *EndNote*: *Capture-Button*, *Citavi*: *Citavi Picker*).



Abb. 4: Vorinstallation des Zotero Web-Importers und Import der Metadaten

Als unübersichtlich, da nicht einheitlich, erweist sich das Netzwerk aus Kooperationen zwischen Datenbanken und Literaturverwaltungsprogrammen. Leider ist nicht immer sofort ersichtlich, welche Software mit welcher Datenbank kooperiert. Hier ist es also nötig, sich die Tools im Einzelnen anzuschauen, bevor sie in den eigenen Forschungsalltag integriert werden.

Vor allem bei umfangreichen Forschungsvorhaben, deren Konzeption eine möglichst vollständige Erfassung der relevanten Forschungsliteratur beinhaltet, profitieren Sie auf der nächsten Stufe eines literaturwissenschaftlichen Forschungsvorhabens von einer digitalen Unterstützung durch die Tools: Organisation und Verwaltung von Forschungsliteratur werden durch die Programme erheblich erleichtert.

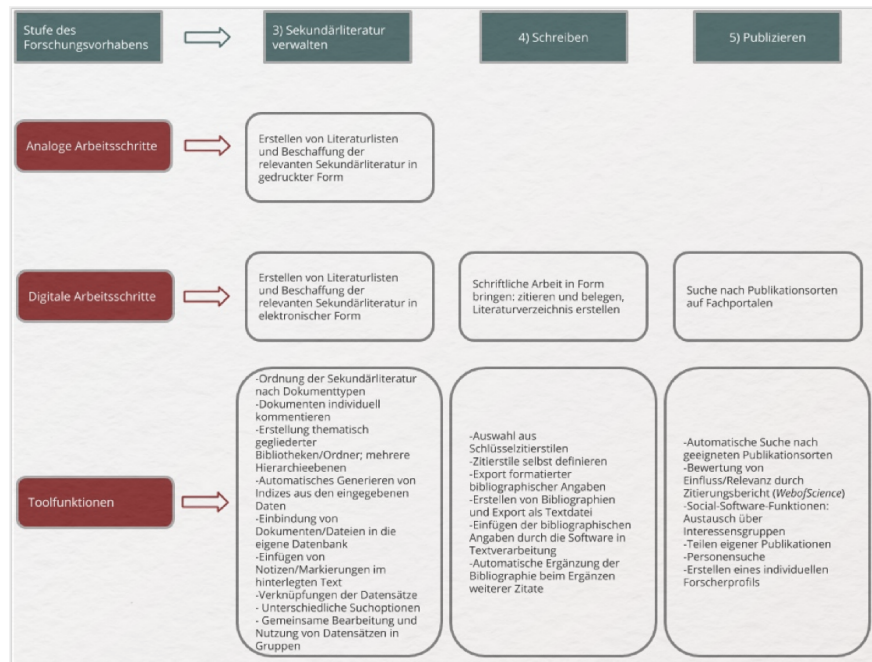


Abb. 5: Systematischer Verlauf eines Forschungsvorhabens: Arbeitsschritte 3, 4 und 5

Dem thematischen Schwerpunkt der Recherche entsprechend, lassen sich digitale Bibliotheken erstellen, in denen Sie die bibliografischen Angaben (Auswahl aus bis zu 50 unterschiedlichen Referenztypen, Titel, Autor*innen, Datum etc.) und importierte Dateien sammeln können. Thematisch verwandte Beiträge aus unterschiedlichen Bibliotheken können miteinander verknüpft und mit Notizen versehen werden. Die Tools ermöglichen das Erstellen einer individuellen Datenbank, die sich stetig erweitern lässt. Ihre Gedanken und Ideen zu Texten können Sie festhalten, jederzeit einsehen und miteinander verbinden. Die Suche nach passenden Zitaten und Quellen bspw. für eine Publikation wird durch den kontinuierlichen Aufbau Ihrer Datenbank und deren thematische Organisation erheblich beschleunigt bzw. erleichtert. Einmal Gelesenes können Sie als bibliografische Angabe festhalten (und evt. zusätzlich als PDF-Dokument anhängen) und mit Schlagworten oder Notizen versehen, die sich über die Suchfunktion (vgl. **Query**) der Tools problemlos wiederfinden lassen. Die sorgfältige Organisation einer individuellen Literaturdatenbank hilft Ihnen nicht nur dabei, Forschungsliteratur zu verzeichnen und thematisch geordnet zu organisieren. Darüber hinaus werden Strukturierung und Abrufbarkeit von Wissen, Gedanken sowie Ideen unterstützt.

Ein weiterer Vorteil gegenüber dem analogen Bibliografieren besteht in der Möglichkeit, Bibliografien kollaborativ zu erstellen und zu verwalten sowie in der Option der Freigabe von Datensätzen für externe Nutzer*innen. An literaturwissenschaftlichen Forschungsprojekten sind häufig mehrere Personen beteiligt. Der Gruppenarbeitsmodus der Tools (mögliche Gruppengrößen variieren tool- und versionsabhängig zwischen drei und 25 Personen) ermöglicht eine effektive Zusammenarbeit, indem relevante Forschungsliteratur von allen Beteiligten zusammengetragen und verwaltet wird. Durch die Sammlung relevanter Forschungsliteratur im literaturwissenschaftlich sachkundigen Kollektiv erstellen Sie auf äußerst effiziente Art und Weise Fachbibliografien, auf die Sie auf der nächsten Stufe des Forschungsvorhabens zurückgreifen können.

Der Schreibprozess stellt einen weiteren Arbeitsschritt dar, den Bibliographietools in unterstützender Funktion begleiten. Aus einer Vielzahl an Zitiertypen (Schlüsselzitiertypen wie MLA, ABA oder Harvard) können Sie den passenden auswählen oder einen eigenen Stil definieren, in dem Bibliografien und Zitate korrekt und einheitlich formatiert werden. Der Zugriff auf besonders hochfrequentierte Referenztypen wie z. B. Fachartikel und deren Zitation wurde um weitere Medientypen wie Multimedia-Anwendungen, TV-Episoden, Social-Media-Beiträge und Diskussionsforen erweitert. Überdies sind die Tools kompatibel mit unterschiedlichen Textverarbeitungsprogrammen (*Citavi*: Word-Add-In; *EndNote*: MS-Word, Apple Pages, OpenOfficeWriter; *Mendeley*: Add-In für MS-Word; *Zotero*: Add-In für MS Word, OpenOffice, LibreOffice, NeoOffice). Dadurch lassen sich bibliografische Angaben in das Textverarbeitungsprogramm einfügen. Neben der Möglichkeit, statische Bibliografien zu exportieren, können Sie dynamische Bibliografien erstellen, die beim Hinzufügen von Literatur automatisch ergänzt werden. Der Vorteil dieser Form des digitalen Bibliografierens besteht vor allem in der einheitlichen und verlässlichen Zitation, die sich problemlos in Dokumente einspeisen und laufend aktualisieren lässt. Vor allem umfangreiche Arbeiten mit entsprechenden Literaturverzeichnissen profitieren an dieser Stelle: Formale Fehler beim Zitieren oder unvollständige Literaturverzeichnisse lassen sich durch den Einsatz eines Literaturverwaltungsprogramms ausschließen.

Die Publikation der Forschungsergebnisse stellt den letzten Schritt dar, bei dem Toolfunktionen wie der *EndNote*

Manuscript-Matcher unterstützen, indem sie eine Liste der am besten für die Veröffentlichung der Arbeit geeigneten Zeitschriften erstellen.

Nicht alle für Literaturwissenschaftler*innen relevanten Fachbibliografien und Kataloge lassen sich jedoch über die Rechercheoption der Tools konsultieren bzw. einige Tools (*Zotero*) verfügen nicht über diese Funktion. Auch für das digitale Bibliografieren gilt ein Grundsatz der Wissenschaftsarbeit: An einem einzigen Ort sind sämtliche für ein Forschungsvorhaben benötigten Informationen nicht zu finden (Cramme und Ritzi 2003, 33). Die durch Softwares zur Auswahl gestellten bibliografischen Datenbanken sollten geprüft und um Datenbanken mit literaturwissenschaftlichen Sammelschwerpunkten ergänzt werden.

5. Technische Grundlagen

Die meisten Datenbanken sind in der Regel innerhalb des Universitätsnetzes kostenlos nutzbar, was durch die in Abbildung 7 abgebildeten Symbole gekennzeichnet wird. Die Kontrolle der Zugangsberechtigung wird durch die Prüfung der **IP-Adresse** durchgeführt. In diesem Fall ist digitales Bibliografieren ortsgebunden. Über den Status der Verfügbarkeit klärt eine einheitliche Zeichengebung auf.

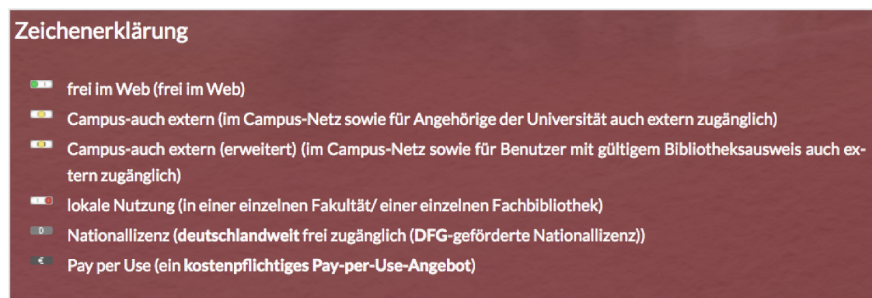


Abb. 7: Status der Verfügbarkeit der Datenbanken (Staats- und Universitätsbibliothek Hamburg)

Seit dem Jahr 2015 sind sämtliche Titeldaten der DNB und die Normdaten der Gemeinsamen Normdatei kostenfrei unter „Creative Commons Zero“-Bedingungen (CC0 1. 0) zur freien Nachnutzung verfügbar. Der Bezug ist über Online-Schnittstellen (Datenshop, SRU- oder OAI-Schnittstelle) nach einmaliger kostenfreier Anmeldung und Autorisierung möglich.

Die digitale Literaturrecherche in Fachbibliografien, die Nutzung der webbasierten Rechercheoption der Tools und die Synchronisation von unterschiedlichen Nutzerkonten, beispielsweise auf unterschiedlichen Rechnern, verlangen in erster Linie nach einer Internetverbindung. Eine äußerst dynamische Softwareentwicklung führt zu der stetigen Weiterentwicklung der Programme. So werden zahlreiche Plug-ins entwickelt, die Sie über die Anbieterseiten herunterladen können.

Wichtig ist zunächst vor allem die Installation des Web-Importers, der den Import der Metadaten aus bibliografischen Datenbanken ermöglicht. Selbst Tools, die eine besonders benutzerfreundliche Oberfläche (vgl. **GUI**) aufweisen, beinhalten eine große Bandbreite unterschiedlicher Funktionen, die zunächst erschlossen werden müssen. Für einige Tools wie *EndNote* existieren Einführungskurse, Selbstlernmaterialien sind in jedem Fall vorhanden. Eine Einarbeitungszeit sollte stets berücksichtigt werden, da die Tools nicht immer intuitiv bedienbar sind. Die Frage danach, welches Tool für die eigenen Belange am besten geeignet ist, lässt sich vor allem über eigenes Ausprobieren beantworten.

Externe und weiterführende Links

- Bibliographie der deutschen Sprach- und Literaturwissenschaft: <https://web.archive.org/web/20220405001633/http://www.bdsl-online.de/BDSL-DB/templates/template.xml?vid=%7B254CC6B9-266C-45F9-99E5-F21E5D1B65D0%7D&contenttype=text/html&Skript=home&lang=de&lang=de> (Letzter Zugriff: 06.11.2024)
- Citavi (Literaturverwaltungsprogramm): <https://web.archive.org/web/20241106121610/https://www.citavi.com/de> (Letzter Zugriff: 06.11.2024)
- Deutsches Textarchiv: <https://web.archive.org/web/20241106121355/https://www.deustextarchiv.de/> (Letzter Zugriff: 06.11.2024)
- Docear (Literaturverwaltungsprogramm): <https://web.archive.org/web/20241106121804/https://docear.org/> (Letzter Zugriff: 06.11.2024)
- EndNote (Literaturverwaltungsprogramm): <https://web.archive.org/web/20241106121837/https://endnote.com/> (Letzter Zugriff: 06.11.2024)
- Fachinformationsdienst Allgemeine und Vergleichende Literaturwissenschaft: <https://web.archive.org/web/20241106122725/https://www.avldigital.de/> (Letzter Zugriff: 06.11.2024)

- Germanistik im Netz: <https://web.archive.org/web/20241106120950/https://www.germanistik-im-netz.de/> (Letzter Zugriff: 06.11.2024)
- Mendeley (Literaturverwaltungsprogramm): <https://web.archive.org/web/20241106122950/https://www.mendeley.com/> (Letzter Zugriff: 06.11.2024)
- TextGrid Repository: <https://web.archive.org/web/20241106113832/https://textgridrep.org/> (Letzter Zugriff: 06.11.2024)
- Zotero (Literaturverwaltungsprogramm): <https://web.archive.org/web/20241106113204/https://www.zotero.org/> (Letzter Zugriff: 06.11.2024)

Bibliographie

- Bartsch, Eberhard. 1979. *Die Bibliographie. Einführung in Benutzung, Herstellung, Geschichte*. München: Sauer.
- . 1989. *Die Bibliographie*. München: Sauer.
- Cramme, Stefan und Christian Ritzi. 2003. Literatur ermitteln. In: *Die Technik wissenschaftlichen Arbeitens. Eine praktische Anleitung*, hg. von Norbert Franck und Joachim Stary, 33–70. Paderborn: Schöningh.
- Domay, Friedrich. 1968. *Formlehre der bibliographischen Ermittlung. Eine Einführung in die Praxis der Literaturschließung*. Stuttgart: Hiersemann.
- Flüh, Marie. 2024b. Ressourcenbeitrag: Germanistik im Netz. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 11. Bibliografie (29. November). doi: 10.48694/fortext.3793, <https://fortext.net/ressourcen/textsammlungen/germanistik-im-netz>.
- . 2024a. Toolbeitrag: Zotero. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 11. Bibliografie (29. November). doi: 10.48694/fortext.3787, <https://fortext.net/tools/tools/zotero>.
- Hansel, Johanna und Lydia Kaiser. 2003. *Literaturrecherche für Germanisten. Studienausgabe*. Berlin: Schmidt.
- Horstmann, Jan. 2024. Ressourcenbeitrag: TextGrid Repository. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 11. Bibliografie (29. November). doi: 10.48694/fortext.3794, <https://fortext.net/ressourcen/textsammlungen/textgrid-repository>.
- Horstmann, Jan und Alexandra Kern. 2024. Ressourcenbeitrag: Deutsches Textarchiv (DTA). Hg. von Evelyn Gius. *forTEXT* 1, Nr. 11. Bibliografie (29. November). doi: 10.48694/fortext.3791, <https://fortext.net/ressourcen/textsammlungen/deutsches-textarchiv-dta>.
- Jeßing, Benedikt. 2010. *Bibliographieren für Literaturwissenschaftler*. Stuttgart: Reclam Verlag.
- . 2017. *Arbeitstechniken des literaturwissenschaftlichen Studiums. Eine Einführung*. Stuttgart: Reclam Verlag.
- Jeßing, Benedikt und Ralph Köhnen. 2012. *Einführung in die neuere deutsche Literaturwissenschaft*. Stuttgart: Metzler.
- Michel, Volker. 2018. Fachinformationsdienst Germanistik. Projektbeschreibung. *Universitätsbibliothek. J.C. Senckenberg UB*. <https://www.ub.uni-frankfurt.de/projekte/fid-germanistik.html> (zugegriffen: 15. April 2019).
- Paschek, Carl. 1999. *Praxis der Literaturinformation Germanistik*. Hg. von Hans-Gert Roloff. Germanistische Lehrbuchsammlung 48. Berlin: Weidler.
- Raabe, Paul. 1980. *Einführung in die Bücherkunde zur deutschen Literaturwissenschaft*. 10. Aufl. Bd. 1. Stuttgart: Metzler.

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.

Commandline Die Commandline (engl. *command line interface* (CLI)), auch Kommandozeile, Konsole, Terminal oder Eingabeaufforderung genannt, ist die direkteste Methode zur Interaktion eines Menschen mit einem Computer. Programme ohne eine grafische Benutzeroberfläche (**GUI**) werden i. d. R. durch Texteingabe in die Commandline gesteuert. Um die Commandline zu öffnen, klicken Sie auf Ihrem Mac „cmd“ + „space“, geben „Terminal“ ein und doppelklicken auf das Suchergebnis. Bei Windows klicken Sie die Windowstaste + „R“, geben „cmd.exe“ ein und klicken Enter.

Feature Unter Features können Einzelfunktionen eines Tools verstanden werden, die beispielsweise komplexe Funktionen wie die Visualisierung eines Textes als **Wordcloud** ermöglichen, oder auch kleinere Funktionseinheiten wie den Abgleich einzelner Spracheigenschaften (**Properties**) mit **annotierten** Beispieltextrn darstellen.

- GUI** GUI steht für *Graphical User Interface* und bezeichnet eine grafische Benutzeroberfläche. Ein GUI ermöglicht es, Tools mithilfe von grafischen Schaltflächen zu bedienen, um somit beispielsweise den Umgang mit der **Commandline** zu umgehen.
- 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.
- IP-Adresse** Die Vernetzung von Computern wird in einem Internetprotokoll (IP) festgehalten, woraufhin jedes angebundene Gerät in diesem Computernetz eine IP-Adresse erhält. So werden die Geräte adressierbar und erreichbar gemacht. Die IP gehört zu den personenbezogenen Daten, da über sie auf Ihre Identität geschlossen werden kann.
- 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 unbekannten Daten verwendet werden.
- 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 Eigennamen, 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.
- OPAC** OPAC steht für *Online Public Access Catalogue* und bezeichnet online zugängliche Bibliothekskataloge.
- PDF** PDF steht für *Portable Document Format*. Es handelt sich um ein plattformunabhängiges Dateiformat, dessen Inhalt auf jedem Gerät und in jedem Programm originalgetreu wiedergegeben wird. PDF-Dateien können Bilddateien (z. B. Scans von Texten) oder computerlesbarer Text sein. Ein lesbares PDF ist entweder ein **OCRter** Scan oder ein am Computer erstellter Text.
- 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**.
- Property** Property steht für „Eigenschaft“, „Komponente“ oder „Attribut“. In der automatischen **Annotation** dienen konkrete Wortheigenschaften wie Groß- und Kleinschreibung zur Klassifizierung von Wörtern oder Phrasen. Durch die Berücksichtigung solcher Eigenschaften in den **Features** eines Tools kann **maschinelles Lernen** bestimmter Phänomene umgesetzt werden. In der manuellen Annotation können als Properties auch Eigenschaften von **Annotationen** benannt werden.
- 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.

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.

URI *Uniform Resource Identifier* (URI) ist ein Identifikator zur eindeutigen Erkennung von Online-Ressourcen wie Webseiten. Im „Raum“ des Internets können so alle Inhalte eindeutig identifiziert werden, unabhängig davon, ob es sich dabei beispielsweise um eine Seite mit Text oder Video handelt. Die am häufigsten verwendete Form eines URI ist die Webseitenadresse, die URL.

Wordcloud Eine *Wordcloud*, oder auch Schlagwortwolke, ist eine Form der Informationsvisualisierung, beispielsweise von Worthäufigkeiten in einem Text oder einer Textsammlung. Dabei werden unterschiedlich gewichtete Wörter, wie die häufigsten Wörter, i.d.R. größer oder auf andere Weise hervorgehoben dargestellt. Die horizontale/vertikale Ausrichtung und die Farbe der dargestellten Wörter hat meistens allerdings keinen semantischen Mehrwert.