



# Tools in den Digital Humanities

---

Vorlesung *Einführung in die Digital Humanities*

Prof. Dr. Christof Schöch  
Wintersemester 2021/22

---



# Sitzungsüberblick

1. Tools in den DH: Was und warum?
2. Literaturverwaltung: Zotero
3. Data Wrangling: OpenRefine
4. Netzwerkanalyse: Cytoscape
5. Korpusanalyse: TXM
6. Fazit

(0) Tools in den DH: Was und warum?

# Tools, sagen Sie? (Definitionsversuch)

- Software: sehr breiter Überbegriff (alles, was nicht Hardware oder Daten sind)
- Spezifischere Begriffe
  - Betriebssystem: Grundlage für den Einsatz von Tools, Vermittlung zur Hardware
  - Programmiersprache: funktional sehr viel offener als ein Tool
  - Tool: Abgrenzbare Anwendungssoftware für einen bestimmten Zweck

# Aspekte der Betrachtung von Tools

- Welche Funktionen hat das Tool?
- Welche Zielgruppe(n) hat das Tool?
- Wer entwickelt das Tool?
- Brauche ich das Tool oder nehme ich lieber Python?
- Um welche Art von Tool handelt es sich?
  - GUI / CLI?
  - Desktop / Webservice?
  - Open Source / proprietär?

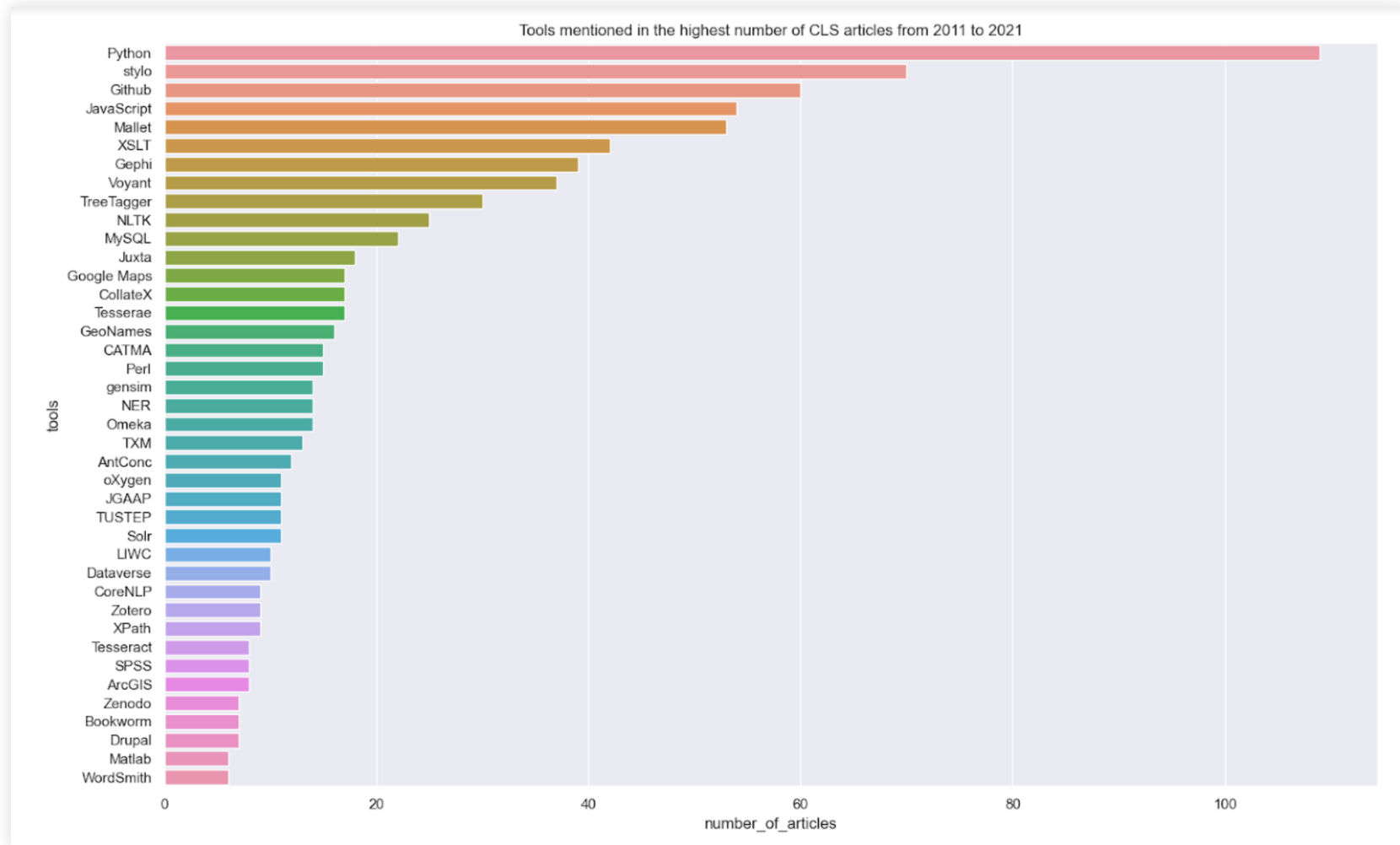
# Arten von Tools

- spezifisch vs. generisch
  - wenige vs. viele Funktionen
  - kleiner vs. großer Teil eines Workflows
  - wenige vs. viele Parameter
  - eine vs. viele Fächer
- forschungsnah vs. publikumsorientiert
  - CLI oder GUI
  - kleiner vs. großer Umfang der Dokumentation
  - niedrige vs. hohe Usability
  - Desktop-Tool oder Webservice

# Tools DH2015-DH2019

1. Python (125)
2. Twitter (82)
3. Gephi (60)
4. JavaScript (59)
5. Omeka (44)
6. GitHub (40)
7. HathiTrust (37)
8. stylo (35)
9. MALLET (33)
10. Google Books (31)
11. Excel (30)
12. MySQL (27)
13. D3.js (23)
14. NLTK (23)
15. WordPress (20)
16. Drupal (19)
17. TextGrid (19)
18. CollateX (18)
19. GeoNames (18)
20. TXM (18)
21. Solr (17)
22. Voyant Tools (17)
23. EEBO-TCP (15)
24. Palladio (15)

# Tools CLS 2011-2021





# (1) Literaturverwaltung: Zotero

# Zotero: Einstieg

- Für die Verwaltung bibliographischer Daten
- Typ: generisch + publikumsorientiert
- Merkmale:
  - Graphische Benutzungsoberfläche
  - Nutzungsfreundlich
  - Umfassende Dokumentation und Forum
  - Fachübergreifend einsetzbar
  - Zielgruppe: Studierende und Wissenschaftler:innen
  - Open Source
- Alternativen: Citavi, Bibtex mit pandoc, Fiduswriter

# Zotero: Funktionen

- Import:
  - aus Katalogen und Datenbanken
  - über ISBN oder DOI
- Verwaltung: Sortieren, Filtern, Schlagworte
- Lektüre: PDF-Reader, Annotation, Notizen
- Integration mit Word Processing (Zitate, Bibliographie)
- Export: Datenexporte und API (z.B. Zotpress)

# Zotero: Screenshots

Desktop

Web

---

Zotpress backend    Zotpress frontend

# Zotero: Ressourcen

- Homepage: <https://zotero.org>
- Screencast aus dem Bachelor:  
<https://www.youtube.com/watch?v=OYG6Fapfueo>
- Einführungsvideo: [https://www.youtube.com/watch?v=PqQp\\_oUUY5w](https://www.youtube.com/watch?v=PqQp_oUUY5w)
- Wikipedia-Artikel: <https://en.wikipedia.org/wiki/Zotero>

## (2) Data Wrangling: OpenRefine

# OpenRefine: Einstieg

- Für die Bearbeitung von tabellarischen Daten
- Typ: generisch + forschungsnah
- Merkmale
  - läuft auf lokalem Server im Browser
  - ein bisschen wie Excel/Calc, aber flexibler
  - fachübergreifend einsetzbar
  - gut dokumentiert
  - laufend weiterentwickelt
  - Open Source
- Alternativen: Python RegEx, BeautifulSoup, Excel/Calc

# OpenRefine: Funktionen

- Exploration: suchen, sortieren, facettieren
- Transformation: korrigieren, verbessern, vereinheitlichen
- Reconciliation: mit Normdaten verbinden und abgleichen
- Vollständige Versionierung



# OpenRefine: Screenshots

View      Filter

---

Datentyp    Facet

# OpenRefine: Ressourcen

- Homepage: <https://openrefine.org/>
- Dokumentation: <https://docs.openrefine.org/>
- Wikipedia-Artikel:  
<https://en.wikipedia.org/wiki/OpenRefine>
- Kurzes Video: <https://www.youtube.com/watch?v=nORS7STbLyk>

# (3) Netzwerkanalyse: Cytoscape

# Cytoscape: Einstieg

- Analyse und Visualisierung von Netzwerkdaten / Graphen
- Typ: forschungnah + generisch
- Merkmale:
  - Desktop-Anwendung mit GUI
  - Großer Funktionsumfang
  - Methoden- aber nicht fachspezifisch
  - Open Source
- Seit 2002, kommt aus der Bioinformatik (!)
- Alternativen: Gephi (DH), NetworkX (Python)

# Cytoscape: Funktionen

- Input: Netzwerkdaten erstellen / importieren
- Analyse: Kennzahlen von Graphen berechnen
- Visualisierung: Netzwerke darstellen

# Cytoscape: Screenshots

Edge-Table	Node-Table
------------	------------

---

Beispiel 1	Beispiel 2
------------	------------

# Cytoscape: Ressourcen

- Homepage: <https://cytoscape.org/>
- Dokumentation: <http://manual.cytoscape.org/>
- Tutorial (Slides): <https://cytoscape.org/cytoscape-tutorials/presentations/network-visualization.html>
- Wikipedia-Artikel: <https://en.wikipedia.org/wiki/Cytoscape>

# (4) Korpusanalyse: TXM



# TXM: Einstieg

- Software für die Analyse von Textkorpora
- Typ: forschungsnah + mittelspezifisch
- Merkmale
  - graphische Benutzungsoberfläche
  - moderat fachspezifisch
  - forschungsnahe Entwicklung
  - Open Source
- Entwickelt seit etwa 2008 in Lyon
- Alternativen: Antconc, WordSmith, stylo, Voyant, CollateX

# TXM: Funktionen

- Import von Texten mit Annotation und Metadaten (TXT, XML, CSV)
- Annotation von Texten (Lemma, POS)
- Suche: Reguläre Ausdrücke auf Wort, Lemma, POS-Ebene
- Analyse: Häufigkeit, KWIC, Kollokationen, Keyness
- Visualisierung: Dokumentlänge, Keyness, Clustering, uvm.

# TXM: Screenshots

Suche nach Mustern    Kollokationen

---

Distinktivität

Clustering

# TXM: Ressourcen

- Homepage: <https://txm.gitpages.humanum.fr/textometrie/en/>
- Dokumentation: <https://txm.gitpages.humanum.fr/textometrie/en/Documentation/>
- Demo der Portal-Version:  
<http://portal.textometrie.org/demo/>
- TXM-Tutorial: <http://christofs.github.io/txm-tutorial/#/>

# Abschluss

# Viele weitere wichtige Tools

- Netzwerkanalyse: Gephi, Cytoscape
- Kartierung: QGIS, Palladio, Geobrowser
- Webdesign: Wordpress, Drupal, Omeka, Github Pages
- Digitale Edition / XML: oXygen und (mit Plugins): Atom, jEdit, VSC
- Forschungsumgebungen: TextGrid Lab, FuD.
- OCR/Transkription: FineReader, OCR4all, Transkribus, Transcribo
- Projektmanagement: Mediawiki, Redmine, Taiga, Kanban uvm.

# Einige nützliche Ressourcen zu Tools

- TAPoR 3: Research Tools for Studying Texts: <https://tapor.ca/pa>
- DH Toolchest von Alan Liu:  
<http://dhresourcesforprojectbuilding.pbworks.com/w/page/69>
- "Which DH Tools Are Actually Used in Research?", <https://welt>

# Literaturhinweise

## Referenzlektüre

- Gibbs, Fred, und Trevor Owens. 2012. „Building better digital humanities tools“. *Digital Humanities Quarterly* 6 (2).  
<http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html>

## Weitere Lektüren

- Bulatovic, Natasa et al. 2016. „Usability in digital humanities-Evaluating user interfaces, infrastructural components and the use of mobile devices during research process“. In *International Conference on Theory and Practice of Digital Libraries*, 335–46. Springer.
- Burghardt, Manuel, und Claudia Müller-Birn. 2019. „Software Engineering in den Digital Humanities“. In: *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik-Informatik für Gesellschaft (Workshop-Beiträge)*. Gesellschaft für Informatik e.V.
- Burghardt, Manuel, und Christian Wolff. 2014. „Humanist-Computer Interaction: Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik“.
- Gold, Nicolas. 2009. „Service-Oriented Software in the Humanities: A Software Engineering Perspective“. *Digital Humanities Quarterly* 3 (4).





Christof Schöch, 2022  
<http://www.christof-schoech.de>

---

Lizenz: Creative Commons Attribution 4.0