

Digitale Werkzeuge zur textbasierten Annotation, Korpusanalyse und Netzwerkanalyse in den Geisteswissenschaften

Digital Philology |

Working Papers in Digital Philology – 02 | 2021

Herausgegeben von

Sabine Bartsch | Evelyn Gius | Marcus Müller | Andrea Rapp | Thomas Weitin

Marcel Frey-Endres (Technische Universität Darmstadt)

Tobias Simon (Hochschule Mainz)



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Impressum

Postadresse:

Technische Universität Darmstadt

Institut für Sprach- und Literaturwissenschaft

Dolivostraße 15

64293 Darmstadt

Website: www.digital-philology.tu-darmstadt.de

Email: sprachli@linglit.tu-darmstadt.de



Zitierhinweis: Frey-Endres, Marcel und Simon, Tobias (2021): Digitale Werkzeuge zur textbasierten Annotation, Korpusanalyse und Netzwerkanalyse in den Geisteswissenschaften. In: Digital Philology | Working Papers in Digital Philology 02|2021. Darmstadt: TUPrints.

Inhaltsverzeichnis

Vorbemerkung	6
1. Einleitung und Überblick	7
2. Grundlage der Auswahl: Verzeichnisse und Übersichten zu digitalen Forschungswerkzeugen für die Geisteswissenschaften	10
3. Übersicht zu den Rahmendaten (technische, anwendungs- und nutzerbezogene Informationen)	11
4. Übersicht zu einzelnen DH-Werkzeugen	15
4.1. Virtuelle Forschungsumgebungen zur Annotation, Analyse, Korpusverwaltung und Edition von Texten	15
4.1.1. e-Humanities Desktop	15
4.1.2. FuD	16
4.1.3. GATE	17
4.1.4. Heurist Data Management	18
4.1.5. TextGrid	20
4.1.6. TUSTEP	22
4.2. Tools zur Textannotation	23
4.2.1. Annotate (A.nnotate)	23
4.2.2. Annotation Studio	24
4.2.3. Annotations	25
4.2.4. Annotator	26
4.2.5. Brat Rapid Annotation	27
4.2.6. Catma	28
4.2.7. Classical Text Editor	29
4.2.8. Coding Analysis Toolkit (CAT) Classical Text Editor	30
4.2.9. Digital Mappa	31
4.2.10. Ediarum	32
4.2.11. eLaborate	34
4.2.12. INCEpTION	35
4.2.13. MorphAdorner	36
4.2.14. Natural Language Toolkit (NLTK)	37
4.2.15. Nota Bene (NB)	38
4.2.16. NowComment	39
4.2.17. PACTE	40
4.2.18. Pundit Annotation	41
4.2.19. QAnnotate	42
4.2.20. Recogito	43
4.2.21. Scripto	44
4.2.22. SPEEDy	45
4.2.23. Text Analysis Markup System (TAMS) Analyzer	46

4.2.24.	Transkribus	48
4.2.25.	UAM CorpusTool	49
4.2.26.	WebLicht	50
4.2.27.	WordFreak	51
4.3.	Text- / Korpusanalytische Tools	52
4.3.1.	ANNIS	52
4.3.2.	AntConc	53
4.3.3.	CorpusExplorer	54
4.3.4.	CQP / Corpus Workbench	55
4.3.5.	Cosmas II (Corpus Search, Management and Analysis System)	56
4.3.6.	Hermetic Word Frequency Finder	57
4.3.7.	Jupyter Notebook	58
4.3.8.	JuxtaCommons	59
4.3.9.	LancsBox (GraphColl)	60
4.3.10.	Rapid Miner	62
4.3.11.	RAW Graphs	63
4.3.12.	Sketch Engine (NoSketch Engine)	64
4.3.13.	Stereoscope	65
4.3.14.	Stylo	66
4.3.15.	TAPORware	67
4.3.16.	Textable	68
4.3.17.	TextDNA	69
4.3.18.	TokenX	70
4.3.19.	Tropes	71
4.3.20.	TXM (Textometrie)	72
4.3.21.	Voyant Tools	74
4.3.22.	Word and Phrase	75
4.3.23.	WORDij	76
4.3.24.	Wordle	77
4.3.25.	Wordtree	78
4.4.	Tools zur Netzwerkvisualisierung	79
4.4.1.	Cytoscape	79
4.4.2.	EgoWeb 2.0	81
4.4.3.	Ezlinavis	82
4.4.4.	Gephi	83
4.4.5.	GraphViz	84
4.4.6.	Histogram	85
4.4.7.	InfraNodus (Texttexture)	86
4.4.8.	Neo4j	87
4.4.9.	Nodegoat	88
4.4.10.	ORA-Lite Netscenes	89
4.4.11.	Pajek	90
4.4.12.	Palladio	91
4.5.	Werkzeuge zur Georeferenzierung und Interaktive Digitale Kartenwerke	92
4.5.1.	ArcGIS	92

4.5.2.	DARE (Digital Atlas of the Roman Empire)	94
4.5.3.	DARIAH Geobrowser	95
4.5.4.	Historypin	96
4.5.5.	Neatline	97
4.5.6.	QGIS	98
5.	Kurzauswertung der DH-Tool-Liste und Fazit	99
6.	Anhang	108
6.1.	Literatur	108
6.2.	Abbildungsnachweise	109

Tabellenverzeichnis

Tabelle 1:	Verzeichnisse digitaler Tools und Ressourcen für die Geisteswissenschaften	11
Tabelle 2:	Heuristische Rahmendaten	12
Tabelle 3:	Schema der Tabellen-Rahmendaten	14

Abbildungsverzeichnis

Abbildung 1:	eHumanities Desktop	15
Abbildung 2:	FuD	16
Abbildung 3:	GATE	18
Abbildung 4:	Heurist Data Management	19
Abbildung 5:	TextGrid	21
Abbildung 6:	TUSTEP	22
Abbildung 7:	Annotate (A.nnotate)	23
Abbildung 8:	Annotation Studio	24
Abbildung 9:	Annotations	25
Abbildung 10:	Annotator	26
Abbildung 11:	Brat Rapid Annotation	27
Abbildung 12:	Catma	28
Abbildung 13:	Classical Text Editor	29
Abbildung 14:	Coding Analysis Toolkit (CAT)	30
Abbildung 15:	Digital Mappa	31
Abbildung 16:	Ediarum	33
Abbildung 17:	eLaborate	34

Abbildung 18: INCEpTION	35
Abbildung 19: MorphAdorner	36
Abbildung 20: Natural Language Toolkit (NLTK)	37
Abbildung 21: Nota Bene (NB)	38
Abbildung 22: NowComment	39
Abbildung 23: PACTE	40
Abbildung 24: Pundit Annotation	41
Abbildung 25: QAnnotate	42
Abbildung 26: Recogito	43
Abbildung 27: Scripto	44
Abbildung 28: SPEEDy	46
Abbildung 29: Text Analysis Markup System (TAMS) Analyzer	47
Abbildung 30: Transcribus	48
Abbildung 31: UAM Corpus Tool	49
Abbildung 32: WebLicht	50
Abbildung 33: WordFreak	51
Abbildung 34: ANNIS	52
Abbildung 35: AntConc	53
Abbildung 36: CorpusExplorer	54
Abbildung 37: CQP / CorpusWorkbench	55
Abbildung 38: Cosmas II (Corpus Search, Management and Analysis System)	57
Abbildung 39: Hermetic Word Frequency Finder	58
Abbildung 40: Jupyter Notebook	59
Abbildung 41: JuxtaCommons	60
Abbildung 42: LancsBox (GraphColl)	61
Abbildung 43: Rapid Miner	62
Abbildung 44: RAW Graphs	63
Abbildung 45: Sketch Engine (NoSketch Engine)	64
Abbildung 46: Stereoscope	65
Abbildung 47: Stylo	66
Abbildung 48: TAPoRware	67

Abbildung 49: Textable	68
Abbildung 50: TextDNA	69
Abbildung 51: TokenX	70
Abbildung 52: Tropes	72
Abbildung 53: TXM (Textometrie)	73
Abbildung 54: Voyant Tools	74
Abbildung 55: Word and Phrase	75
Abbildung 56: WORDij	77
Abbildung 57: Wordle	78
Abbildung 58: Wordtree	79
Abbildung 59: Cytoscape	80
Abbildung 60: EgoWeb 2.0	81
Abbildung 61: Ezlinavis	82
Abbildung 62: Gephi	83
Abbildung 63: GraphViz	84
Abbildung 64: Histogramph	85
Abbildung 65: InfraNodus (Texttexture)	86
Abbildung 66: Neo4J	87
Abbildung 67: Nodegoat	88
Abbildung 68: ORA Lite Netscenes	89
Abbildung 69: Pajek	90
Abbildung 70: Palladio	91
Abbildung 71: ArcGis	93
Abbildung 72: DARE (Digital Atlas of the Roman Empire)	94
Abbildung 73: DARIAH Geobrowser	95
Abbildung 74: Historypin	96
Abbildung 75: Neatline	97
Abbildung 76: QGIS	98

Vorbemerkung

Die vorliegende Publikation ist aus dem Projekt *Humanist Computer Interaction auf dem Prüfstand* hervorgegangen, einem Verbundvorhaben des Fachgebiets Germanistik - Computerphilologie und Mediävistik der Technischen Universität Darmstadt, der Forschungsgruppe WIMM am Fachbereich Wirtschaft der Hochschule Mainz und dem Arbeitsbereich Alte Geschichte der Johannes Gutenberg-Universität Mainz (Verbundkoordination), das im Rahmen der VIP+ Initiative des BMBF im Zeitraum vom 01.10.2017 – 30.10.2020 gefördert wurde.

Sämtliche Textbausteine, Übersichten und Listen zu digitalen Werkzeugen sind an der Technischen Universität Darmstadt im Zuge intensiver Recherche- und Evaluationsprozesse zu den Angeboten für geisteswissenschaftliche Texterschließung und –analyse in der gegenwärtigen digitalen Werkzeuglandschaft entstanden. Die Hochschule Mainz war begleitend in die Entstehung der Publikation eingebunden und trägt insbesondere durch Visualisierung der tabellarisch aufgeführten funktionellen und methodischen Gesichtspunkte der einzelnen digitalen Werkzeuge im Rahmen der Entwicklung einer *Digital Research Tool Landscape* zur Erhöhung der Sichtbarkeit des Ergebnisses bei. [Simon, Tobias; Frey-Endres, Marcel; Rossner, Alexander; Pagel, Sven (2020). *Digital Humanities Research Tool Landscape* (Version v1.0.0). <http://doi.org/10.5281/zenodo.3922123>].

Eine weitere Veröffentlichung aus der Verbundkollaboration bildet ein Kurztext, der eine Synopse aus wesentlichen Bausteinen dieses Arbeitspapiers und den Forschungsfragen und Zielen im Kontext der *Technology Landscape* beinhaltet [Simon, Tobias; Frey-Endres, Marcel; Rossner, Alexander; Pagel, Sven (in Vorbereitung). *Digital Humanities Research Tool Landscape – Ein Systematisierungsinstrument zur Auswahl von digitalen Werkzeugen in den Geisteswissenschaften*].

1. Einleitung und Überblick

Mit der zunehmenden Digitalisierung von Arbeitsformen, Forschungsprozessen und Wissensgebieten in den Geisteswissenschaften geht auch eine Ausdifferenzierung entsprechender Aufbereitungs-, Erschließungs-, Analyse- und Edierungswerkzeuge einher, deren bloße Fülle sich nur noch schwerlich überblicken oder gar nach einem einheitlichen Prinzip systematisieren lässt. Das Spektrum der digitalen Tool-Landschaft lässt sich jedoch grob nach verschiedenen Gesichtspunkten einteilen.

Ein erstes Kriterium ist die methodische Breite der Anwendungen. Einerseits findet sich eine ganze Reihe umfangreicher virtueller Forschungsumgebungen, die darauf abzielen, einen großen Teil des geisteswissenschaftlichen Forschungsprozesses von der Datenaufnahme bis zur Produktion digitaler Editionen abzubilden. Andererseits existieren unzählige Pilot- und Spezialentwicklungen, die auf besondere Materialien, Problemdomänen und Forschungsfragen zugeschnitten sind. Einen zweiten Gesichtspunkt bildet die Gewichtung von Effizienz und methodischer Flexibilität. Digitale Anwendungen können entweder stark auf ein effizientes und zeitsparendes Prozessieren der digitalen Erschließungsstrategien ausgerichtet sein oder räumen der Gestaltung der Datenannotation und den daraus resultierenden Analysemöglichkeiten möglichst große Spielräume ein. Ein drittes Merkmal besteht in der unterschiedlichen Komplexität der Anwendungen, die sich aus dem Entwicklungshintergrund und den jeweiligen Zielgruppen, für die ein Werkzeug entwickelt wurde, ergibt. Die Bandbreite reicht von einfachen und ohne weitere Kenntnisse intuitiv nutzbaren Tools bis zu sehr komplex aufgebauten oder nur mit entsprechenden Kenntnissen von Markup- oder Skriptsprachen bedienbaren Lösungen. Viertens lassen sich digitale Werkzeuge zur Textanalyse nach methodischem Anspruch differenzieren. So finden sich neben zahlreichen auf textstatistische und quantitative Auswertung großer Korpora ausgerichteten Werkzeugen auch zunehmend Anwendungen, die ihren Fokus stärker auf die Integration qualitativer geisteswissenschaftlicher Verfahren legen.

Die Gründe für diese Heterogenität sind ebenso vielfältig. Die Geisteswissenschaften sind durch eine besonders große methodische Breite geprägt, die es durch entsprechende Werkzeuge abzubilden gilt (vgl. Kaden 2016). Diese Vielfalt wird vor allem dann zur Herausforderung, wenn interoperable Lösungen für verschiedene methodische Konzepte und geisteswissenschaftliche Daten mit dem Anspruch einer Vereinheitlichung und Standardisierung von digitalen Verarbeitungsprozessen und Datenformaten in Einklang gebracht werden sollen. Die Entscheidung, welche Daten aus welchen Gründen und in welcher Weise erhoben, organisiert, systematisiert, kategorisiert werden sollen, kann einmalig zu

Beginn des Forschungsprogramms oder im Forschungsprozess auch immer wieder neu zu fällen sein. Dementsprechend existiert ein Bedarf an hocheffizienten, aber wenig modifizierbaren „Flaschenhals“-Lösungen und andererseits an sehr flexiblen, aber mit höherem Aufwand zu nutzenden Werkzeugen. Als relativ junges wissenschaftliches Arbeitsfeld stehen die Digital Humanities zudem vor der Aufgabe, sich sowohl über ihre disziplinäre Einordnung zwischen Geistes- und Informationswissenschaften zu vergewissern als auch mit Fragen der methodologischen Verknüpfung hermeneutischer und algorithmischer Verfahren auseinanderzusetzen (Baum & Stäcker 2015, Kaden 2016). Die experimentelle Vielfalt und Breite der digitalen Werkzeuge unterstreicht diesen disziplinären Sondierungsprozess. Nicht zuletzt weisen die Digital Humanities durch ihre transdisziplinäre Charakteristik zwischen Geisteswissenschaften und Informatik ein erweitertes Spektrum an Aufgabenbereichen zwischen Anwendungsentwicklung und methodologischer Arbeit auf (vgl. Nerbonne 2015). Einerseits sollen große Datenmengen quantitativ und automatisiert erschließbar sowie quantitativ ausgewertet werden können, andererseits verlieren die „klassischen“ textwissenschaftlichen Prozesse hermeneutischer Analyse und theoriegeleiteter Interpretation keineswegs an Bedeutung, sondern stehen im Sinne einer Methodologie der „Datenhermeneutik“ gerade bei der Erschließung und Analyse von Texten vor wichtigen Herausforderungen (Gius & Jacke 2015). Beispiele hierfür sind die Modellierung, Analyse und Interpretation komplexer semantischer und linguistischer Relationen wie z.B. in der Diskursanalyse (vgl. Bender 2020, Gredel 2018) oder auch das Problem der Datierung von Unsicherheiten, Auslegungsvarianten und Zweifeln in Bezug auf Quellenmaterialien wie in den Geschichtswissenschaften (vgl. Kuczera et al. 2019). Angesichts des weiten Spektrums an digitalen Werkzeugen wird neben der Problematik der informationstechnischen Implementierung geisteswissenschaftlicher Methoden auch die Bedeutung von Konzepten zur Evaluation der *Usability* von Benutzeroberflächen und eines an der geisteswissenschaftlichen NutzerInnen-Gruppe ausgerichteten *Usability Engineerings* bei der Realisierung digitaler Werkzeuge immer deutlicher. Unterstützung bei der Auswahl als auch der Produktion geeigneter Anwendungen können hierbei Systematiken zur Identifikation und Integration spezifischer Patterns in digitalen Werkzeugen bieten, wie dies beispielsweise für Werkzeuge zur linguistischen Annotation gezeigt werden konnte (Burghardt 2014).

Die hier nur in Umrissen skizzierten Problemfelder bei der Wahl der „passenden“ Tools beschreiben zugleich die Herausforderungen, die im BMBF-Projekt „Humanist Computer Interaction auf dem Prüfstand“ von Beginn an deutlich wurden. Zielsetzung des Projektes war es, die Innovationspotentiale digitaler Methoden in den Geisteswissenschaften auf Grundlage der Forschung an einem exemplarischen spätantiken Briefkorpus, den *Variae* des gotischen

Hofbeamten Cassiodor, in digitalen Workshops mit FachwissenschaftlerInnen anhand geeigneter digitaler Werkzeuge zu validieren. Der anfänglich geplante und überwiegende Einsatz der Forschungsumgebung *TextGrid* erwies sich jedoch aus mehreren Gründen als nicht zielführend. Zum Ersten ließ sich der hierarchisch aufgebaute und stark an linguistischen und philologischen Kriterien entwickelte TEI/XML-Standard nur bedingt mit dem Ziel einer multidimensionalen historisch-semanticen und philologischen Annotation und Kommentierung der *Variae* verbinden. Zum Zweiten stand das Ziel, in den Workshops an einer kritischen Masse bereits annotierter Daten verschiedene digitale Analysemethoden zu vermitteln und zu erproben, im Konflikt mit den Möglichkeiten, anhand der überwiegend auf manuelle und halbautomatische XML-Codierung ausgerichteten *TextGrid*-Umgebung in kurzer Zeit einen nutzbaren Datenbestand aufzubauen. Zum Dritten sollten im Sinne eines Querschnitts neben Annotations- und Editionsarbeiten auch analytische Methoden wie korpuslinguistische und netzwerkanalytische Zugänge vermittelt werden, so dass sich die Nutzung einzelner modularer Tools für die Gestaltung der Validierungsworkshops eher empfahl als der Einsatz einer einzelnen monolithischen Anwendung. Das Projekt stand vor der Aufgabe, aus dem umfangreichen Angebot an digitalen Werkzeugen angemessene Lösungen herauszufinden, die eine effektive Datenerschließung und -analyse erlauben. Zugleich sollten sich diese Lösungen für den aktiven Einsatz zur Vermittlung grundlegender digitaler Methoden in Workshops mit GeisteswissenschaftlerInnen eignen, die bislang nur in geringem Maße digital gearbeitet hatten.

Die im Folgenden zusammengestellte Auflistung digitaler Werkzeuge für die Geisteswissenschaften ist das Ergebnis dieser Recherche. Sie beinhaltet Steckbriefe von digitalen Anwendungen zur Annotation, Korpusanalyse, Netzwerkanalyse und Georeferenzierung geisteswissenschaftlicher Textdaten. Sie deckt somit einen in zweierlei Hinsicht ausschnitthaften Bereich aus dem Gesamtspektrum digitaler Werkzeuge für die Geisteswissenschaften ab, der sich aus den methodischen Kerntätigkeiten im Humanist-Projekt ergeben hat. Zum einen liegt der Fokus ausschließlich auf digitalen Anwendungen zur geisteswissenschaftlichen Erschließung und Untersuchung von *Texten*. Zum anderen wird mit dem Schwerpunkt auf Tools zur *(Text-)Annotation* und *(Text-)Datenanalyse* zwar ein spezifisches, aber doch wesentliches Segment des gesamten digitalen Forschungsprozesses von der Erfassung und Vorbereitung des Datenmaterials bis zur Veröffentlichung und Archivierung nachnutzbarer Resultate abgebildet. Die zusammengetragenen digitalen Werkzeuge wurden, soweit dies im Rahmen der Recherche möglich war, installiert bzw. aktiviert sowie ihre Funktionen anhand von Beispieldaten getestet. Alternativ bzw. ergänzend wurden die Anwendungsgrundlagen anhand der in den meisten Fällen vorliegenden Dokumentationen

geprüft. Die Kurzbeschreibungen fassen die wesentlichen Eigenschaften der digitalen Tools auf Grundlage der Anwendungstläufe und wichtigsten nutzerrelevanten Informationen auf den Dokumentationsseiten zusammen.

Die vorliegende Liste soll insbesondere Forschenden, die bislang wenig mit digitalen Methoden in den Geisteswissenschaften arbeiten, eine Orientierungshilfe in der Werkzeug-Landschaft geben. Sie erhebt weder Anspruch auf auch nur annähernde Vollständigkeit noch auf erschöpfende Darstellung der Bandbreite von Anwendungskonzepten in den ausgewählten Methodenbereichen. Inhaltlich ist sie das Ergebnis einer kontinuierlichen Recherche auf Basis der im Rahmen des Humanist-Projektes angestellten Erwägungen und Anforderungen an digitale Werkzeuge. Überdies handelt es sich um eine Momentaufnahme des gegenwärtigen Angebotes an digitalen Forschungslösungen, die in Abhängigkeit von neuen Entwicklungen und nicht mehr unterstützten oder abrufbaren Anwendungen in Zukunft zumindest partiell wieder an Aktualität verlieren dürfte.

2. Grundlage der Auswahl: Verzeichnisse und Übersichten zu digitalen Forschungswerkzeugen für die Geisteswissenschaften

Als Grundlage der Recherche digitaler Tools zur Annotation, Korpusanalyse sowie zur Netzwerk- und Geoanalyse von Textdaten dienen im Verlauf der Zusammenstellung eine Reihe einschlägiger Listen und Verzeichnisse digitaler Tools, die online abrufbar sind. Die drei umfangreichsten Übersichten bieten das TAPoR Gateway der University of Alberta, das Digital Toychest for the Humanities der University of California sowie die Community-basierte DIRT Directory, die gegenwärtig allerdings nur noch über Wayback abrufbar ist. Daneben wurden weitere Listen und Übersichten verschiedener Verbundinitiativen und Institute gesichtet und zuletzt auch gezielt neuere Anwendungen recherchiert, die bislang noch nicht in Zusammenstellungen gelistet wurden.

Name des Verzeichnisses	Beschreibung
DIRT Directory	Großes Verzeichnis von DH-Werkzeugen
Digital Toychest for Humanities	Übersicht über verschiedene DH-Methoden und Werkzeuge
TAPoR – Digital Text Research Tools	Umfassendes Recherche-Gateway DH-Werkzeugen
History Online	Großes Search Engine für DHistory Tools
DMI Tools	Werkzeugliste der Digital Humanities Initiative
DARIAH-DE Werkzeuge und Dienste	Werkzeug- und Dienstliste technischer Angebote bei DARIAH-
Carolina Digital Humanities Initiative	Liste von DHumanities Tools
D4 Virtual Research Environments	Verzeichnis virtueller Forschungsumgebungen

ZMI Tool-Liste	Liste des Zentrums für Medien und Interaktivität der Universität Gießen
Digital Toolbox for Historians	Tool-Liste für Historiker (Pinterest-Eintrag)
DH-Werkzeuge für nicht-lateinische Daten	DH-Tool-Liste der FU Berlin
Digital Textuality Ressources Pages	Liste von Tools zur Aufbereitung von Medien
Text Visualisation Techniques Browser	Übersicht der Linnäus University zu Visualisierungswerkzeugen
Visual Complexity	Liste zu Visualisierungstechniken der Parsons School New York
Purdue Toolbox	Verzeichnis freier Open-Source-Software von Drittanbietern

Tabelle 1: Verzeichnisse digitaler Tools und Ressourcen für die Geisteswissenschaften

3. Übersicht zu den Rahmendaten (technische, anwendungs- und nutzerbezogene Informationen)

Neben den Kurzbeschreibungen und beispielhaften Abbildungen der digitalen Werkzeuge enthält die Auflistung tabellarische Übersichten mit verschiedenen Rahmendaten, die über verschiedene *technische Aspekte* wie den Anwendungstyp, die Art der Aktivierung, unterstützte Betriebssysteme bzw. benötigte andere Anwendungen, den Lizenztyp sowie die Art der Dateneingabe und die bei der Speicherung sowie bei Im- und Export unterstützten Datenformate informiert.

Des Weiteren geben die Übersichten Auskunft über die mit der jeweiligen Anwendung durchführbaren *geisteswissenschaftlichen Methoden* im Bereich der Annotation und Analyse von Textdaten. Ein nicht geringer Teil der Werkzeuge deckt sowohl Annotationsaufgaben als auch Analysemethoden ab, so dass eine eindeutige Zuordnung zu den vier Bereichen Annotation, Korpusanalyse, Netzwerkanalyse und Georeferenzierung nicht immer möglich ist. In solchen Fällen wurden die Anwendungen entsprechend ihres Hauptzweckes rubriziert. Ergänzend werden außerdem Angaben zu bestimmten computertechnischen oder fachbezogenen Vorkenntnissen gemacht, die für eine Nutzung der Anwendung erforderlich oder hilfreich sind.

Ein dritter Bereich enthält Angaben zu nutzerbezogenen Faktoren. Grundlage für diese Kriterien sind die in der NutzerInnenforschung bewährten *10 Usability Heuristics for User Interface Design* (Nielsen 1995), die in die vier Teileigenschaften Semantik des Informationsdesigns, Funktionsumfang, Struktur der Steuerelemente und Bedienbarkeit zusammengeführt wurden. Zweck dieser Anpassung der Heuristik ist es, keine Evaluation oder Bewertung abzugeben, sondern die Tools in einem Kontinuum zwischen eher übersichtlichen, leicht zu bedienenden Anwendungen mit wenigen Funktionen und eher komplexen, leistungsfähigen und in ihrer Bedienung anspruchsvolleren Werkzeugen zu verorten.

Heuristik nach Nielsen (1994)	Einordnungskriterien Liste
1. Fehlervermeidung	NutzerInnen-Unterstützung
2. Fehlererkennung	Dokumentation
3. Benutzungskontrolle und -freiheit	Integrierte Eingabe- / Funktionshilfe
4. Hilfe und Dokumentation	Automatische Fehlerrückmeldung
1. Vereinbarkeit des Systems mit der Wirklichkeit	Konzeptionelle Gesichtspunkte der Anwendung
2. Konsistenz und Standards	Semantik des Informationsdesigns
3. Wiedererkennung statt Erinnerung	Funktionsumfang
4. Sichtbarkeit des Systemstatus	Struktur der Steuerelemente
5. Flexibilität und Effizienz der Nutzung	Bedienung
6. Ästhetik und minimalistisches Design	

Tabelle 2: Heuristische Rahmendaten

Statt in Form von Bewertungskriterien werden diese heuristischen Prinzipien in neutralisierter Form zu grundlegenden Eigenschaften des Programms umformuliert und in zwei Bereiche gegliedert.

Zum einen werden verschiedene Formen der NutzerInnenunterstützung aufgeführt, die im Rahmen der Werkzeuge angeboten werden, wie den Umfang der Dokumentation, integrierte Eingabehilfen, die Kommunikation von Ein- und Ausgabebefehlern sowie Supportangebote. Zum anderen werden für interessierte NutzerInnen belangreiche konzeptionelle Aspekte der Anwendung mittels einer wertungsfreien Terminologie beschrieben, um eigene Erwartungen an das Programm und seine individuelle Usability besser einschätzen zu können. So wird beispielsweise unter der Kategorie „Semantik des Informationsdesigns“ eine Aussage darüber getroffen, ob die Anwendungs- bzw. NutzerInnenkommunikation eher an der technischen Funktion selbst oder eher an der – hier: geisteswissenschaftlichen – Domäne orientiert ist, ohne dieses Verhältnis als mehr oder weniger nutzerfreundlich zu bewerten.

Im Folgenden ist das Schema der Rahmendaten nochmals detailliert beschrieben:

Technische Grundlagen		
• Grundlegende Informationen zu der technischen Beschaffenheit der Anwendung		
Art der Anwendung	beschreibt die Art der technischen Implementierung in bestehende Systeme	<ul style="list-style-type: none"> • Portable Software • Desktopanwendung (lokal) • Desktopanwendung (remote) • Browserapplikation (lokale Server) • Website (externer Server)
Einrichtung	beschreibt die erforderlichen Maßnahmen, die zur vollständigen Inbetriebnahme der Anwendung benötigt werden	<ul style="list-style-type: none"> • Installation • Ausführbare EXE-Datei • Ausführbare JAR-Datei • Server Setup • Registrierung

BS	Betriebssysteme, in denen die Anwendung genutzt werden kann	<ul style="list-style-type: none"> • Win • Mac OS • Linux • Andere
NutzerInnen-Unterstützung <ul style="list-style-type: none"> • Dokumentation von Entwicklung, Funktionsweise und Quellcode • Anwendungsinterne Unterstützung bei der Bedienung • Sonstige Unterstützung (Support) 		
Dokumentation	<p>vollständig und detailliert</p> <p>unvollständig, aber detailliert</p> <p>grundlegende Dokumentation</p> <p>keine Dokumentation</p>	<ul style="list-style-type: none"> • ausführlich • elementar • begrenzt • keine
Integrierte Eingabe- / Funktionshilfe	Die Anwendung unterstützt durch eine informative NutzerInnenführung bei allen wesentlichen Funktionen, nur bei Grundfunktionen oder gar nicht	<ul style="list-style-type: none"> • elementar • begrenzt • nein
Automatische Fehlerrückmeldung	Die Anwendung kommuniziert Fehler bei der Steuerung und Eingabe oder Fehler werden nicht kommuniziert	<ul style="list-style-type: none"> • ja • nein
Support	Es existieren zusätzliche Optionen der NutzerInnenunterstützung	<ul style="list-style-type: none"> • umfassend (direkter Support) • elementar (Foren, Issue Tracking) • begrenzt (FAQs, Tutorials) • nein
Zwecke und benötigte NutzerInnenkenntnisse <ul style="list-style-type: none"> • Unterstützte digitale Methoden und Arbeitsprozesse • Besondere erforderliche Kenntnisse zur Nutzung der Anwendung 		
Annotation	Prozesse der deskriptiven und analytischen / linguistischen und semantischen Erschließung des Datenmaterials	<ul style="list-style-type: none"> • manuelle Annotation (mAnn) • (teil-)automatisierte Annotation (aAnn) • Kategorisierung (Kat) • Schema-Erstellung (Schem) • linguistische Annotation (lAnn) • semantische Annotation (sAnn)
Korpusanalyse	Prozesse der quantitativen (textstatistischen) und qualitativen (texthermeneutischen) Analyse von Eigenschaften und Attributen in Texten und Korpora	<ul style="list-style-type: none"> • Linguistische Analyse (lingA) • Semantische Analyse (semA) • Diskursanalyse (discA) • Topical Mapping (toMap) • Sentimentanalyse (sentA) • Visualisierung (vis)
Netzwerkanalyse	Analyse und Visualisierung relationaler Eigenschaften, Strukturen und Prozesse in Texten und Korpora	<ul style="list-style-type: none"> • Netzwerkanalyse (netwA) • Graphenmodellierung (GMod) • Netzwerkvisualisierung (vis) • Zeitreihen-Analyse (timA)

Georeferenzierung	Analyse und Visualisierung geographischer / räumlicher Beziehungen und Prozesse von Texten und Korpora	<ul style="list-style-type: none"> • Geoanalyse (geoA) • Geovisualisierung (vis) • Zeitreihen-Analyse (timA) • Kartenerstellung (map)
Benötigte NutzerInnenkenntnisse	besondere technische oder fachliche Kenntnisse, die zur Anwendung erforderlich und hilfreich sind	<ul style="list-style-type: none"> • Programmierkenntnisse • Markup-Sprachen (z.B. XML) • Graphentechnologien • Computerlinguistische Kenntnisse
Eingabemodi und Dateiformate		
Steuerungs- und Eingabemodi	Steuerung: WIMP (Windows, Icons, Menus, Pointer) oder Kommandozeile Dateneingabe: Reintext oder Markup	<ul style="list-style-type: none"> • WIMP • Befehlszeile • Reintext • Markup
Dateiformate	Von Anwendung (lokal, extern) gespeicherte Formate	
Import	Formate, die in die Anwendung importiert werden können	
Export	Formate, die aus der Anwendung exportiert werden können	
Konzeptionelle Gesichtspunkte der Anwendung		
Semantik des Informationsdesigns	Anwendung kann eher auf technisch-funktionale oder eher auf domänenspezifische (hier geisteswissenschaftliche) Semantiken orientiert sein	<ul style="list-style-type: none"> • funktionsorientiert • domänenorientiert
Funktionsumfang	Anwendung kann ein sehr umfassendes Spektrum, einen substanziellen Bestand oder einen geringen Umfang an Funktionen anbieten	<ul style="list-style-type: none"> • hoch • moderat • gering
Struktur der Steuerelemente	Gliederung der Steuer- und Eingabeelemente kann, z.T. abhängig von Semantik und Funktionsumfang, einfach, übersichtlich oder komplex strukturiert sein	<ul style="list-style-type: none"> • einfach • übersichtlich • differenziert • komplex
Bedienung	Bedienung der Anwendung kann, abhängig von Semantik, Funktionsumfang und Steuerungsstruktur, eher leicht, moderat, anspruchsvoll oder schwierig sein	<ul style="list-style-type: none"> • leicht • moderat • anspruchsvoll • schwierig

Tabelle 3: Schema der Tabellen-Rahmendaten

4. Übersicht zu einzelnen DH-Werkzeugen

4.1. Virtuelle Forschungsumgebungen zur Annotation, Analyse, Korpusverwaltung und Edition von Texten

4.1.1. e-Humanities Desktop

eHumanities Desktop ist eine Onlineplattform zur Unterstützung der Modellierung, Verwaltung, Abfrage und Bearbeitung multimedialer geisteswissenschaftlicher Daten. Es bietet eine intuitiv bedienbare Umgebung ähnlich dem Windows-Desktop, über die sich geisteswissenschaftliche Daten speichern, zuweisen, miteinander verknüpfen und gemeinsam weiterbearbeiten lassen. Auf Grundlage der serviceorientierten Architektur zur effizienten kollaborativen Ressourcenverwaltung bietet die Forschungsumgebung verschiedene Werkzeuge zur Bearbeitung und Analyse von Dokumenten an, angefangen von der korpuslinguistischen und semantischen Erschließung in TEI/XML über die Verwaltung lexikalischer Ressourcen und der Klassifikation von Texten und Bilddaten bis hin zur Durchsuchbarkeit, Analyse und Visualisierung sprachlicher Netzwerke.

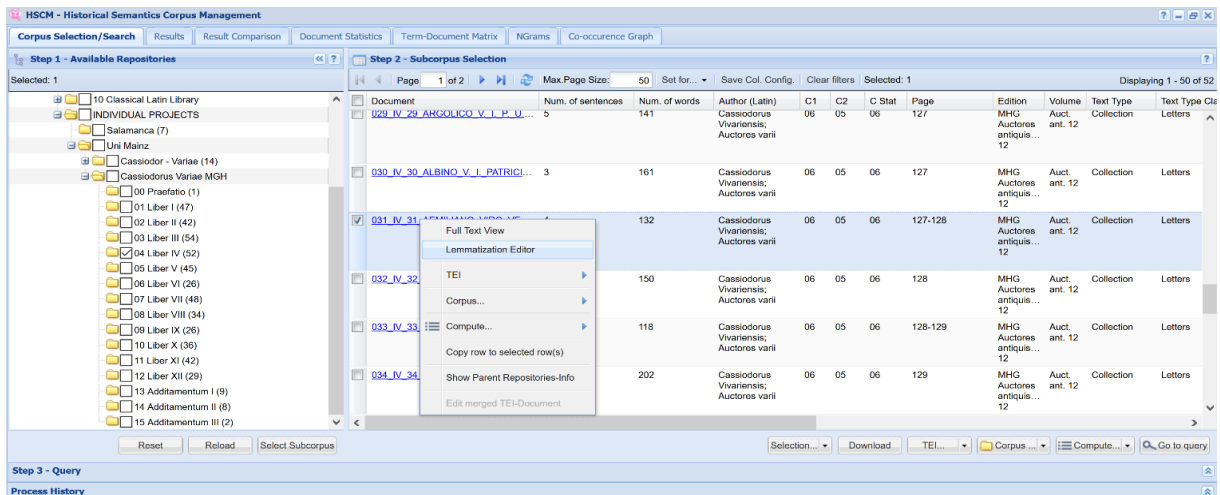


Abbildung 1: eHumanities Desktop

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizentyp	Website Registrierung
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	elementar elementar ja begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation

Einzelmethoden	mAnn, Kat, Schem, lAnn, sAnn, netwA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	tei, xml
Import	tei, xml
Export	tei, xml
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://www.texttechnologylab.org/applications/ehumanities-desktop/

4.1.2. FuD

FuD ist eine virtuelle Forschungsumgebung für die geistes- und sozialwissenschaftliche Forschung. Die modular aufgebaute Anwendung bildet den kompletten Forschungsprozess von der Datensammlung, -annotation, -analyse und -aufbereitung bis hin zur Publikation und Archivierung ab und unterstützt die zeit- und ortsunabhängige kollaborative Forschungsarbeit. FuD kann zur Erstellung digitaler Werkausgaben sowie Quellen- und Regesteneditionen, der Erschließung von Archivbeständen und kollaborativen Organisation und Pflege von Materialsammlungen, zum Aufbau von Personendatenbanken und zur Durchführung von Diskurs- und Netzwerkanalysen genutzt werden.

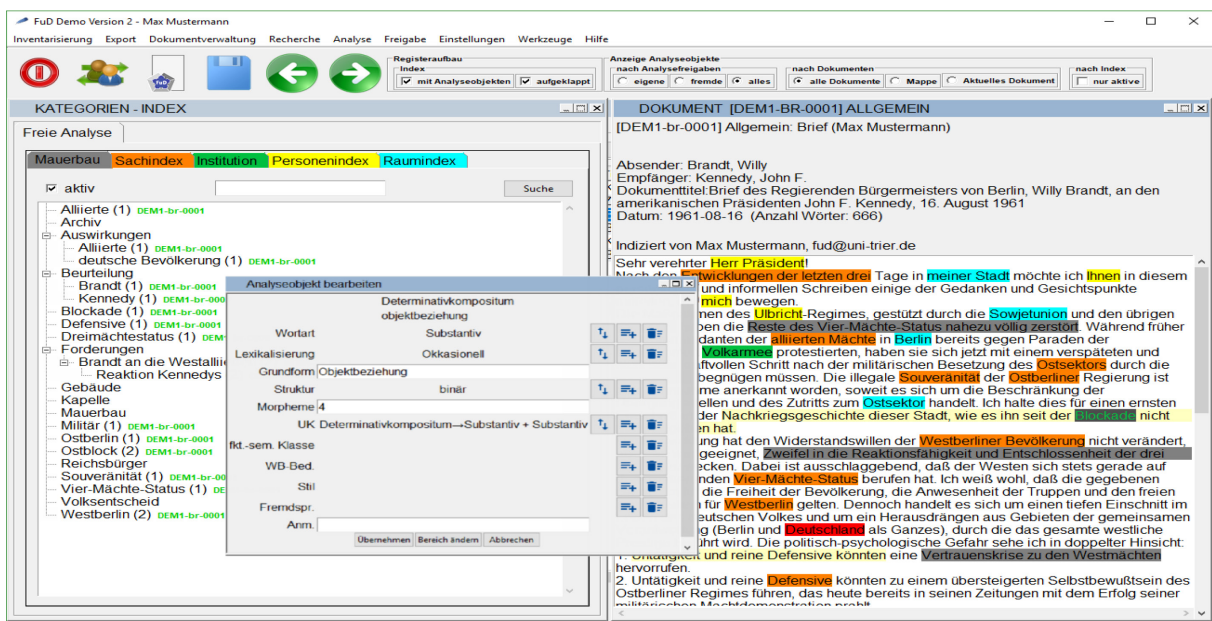


Abbildung 2: FuD

Technische Daten	
Art der Anwendung	Desktopanwendung (remote)
Aktivierung	Registrierung, Server Setup
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source, Payware, Freeware

Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Schem, Kat, lAnn, sAnn, lingA, semA, discA, netWA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	tei, xml, rtf
Import	
Export	rtf, tei, xml, txt, LaTeX
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://fud.uni-trier.de/

4.1.3. GATE

GATE ist eine umfangreiche JAVA-basierte Werkzeugsuite, die sich aus verschiedenen Softwarekomponenten und Webservices zur Verarbeitung aller Formen menschlicher Sprache sowie zur Analyse von Texten aller Formen und Größen zusammensetzt. Die Plattform ist eine der umfangreichsten Open-Source-Entwicklungen im Bereich der Sprachverarbeitung und bietet zahlreiche Möglichkeiten für computerlinguistische Analysen, Natural Language Processing und Language Engineering. Die GATE-Werkzeugfamilie beinhaltet verschiedene Kernkomponenten und arbeitet auf Grundlage des Informationsextraktionssystems ANNIE, das Module zur Tokenisierung, POS-Tagging, Named Entity Erkennung und Koreferenzierung enthält. Die GATE Cloud erlaubt gehostete Textverarbeitung in großem Maßstab und enthält zahlreiche Tools zur automatisierten Annotation von Texten. Die Webanwendung GATE Teamware eignet sich für kollaborative semantische Annotationsprojekte. Die Desktop-Anwendung GATE Developer enthält zahlreiche interaktive Werkzeuge zur Komposition von Verarbeitungsprozessen zur Annotation und Analyse von Textdokumenten.

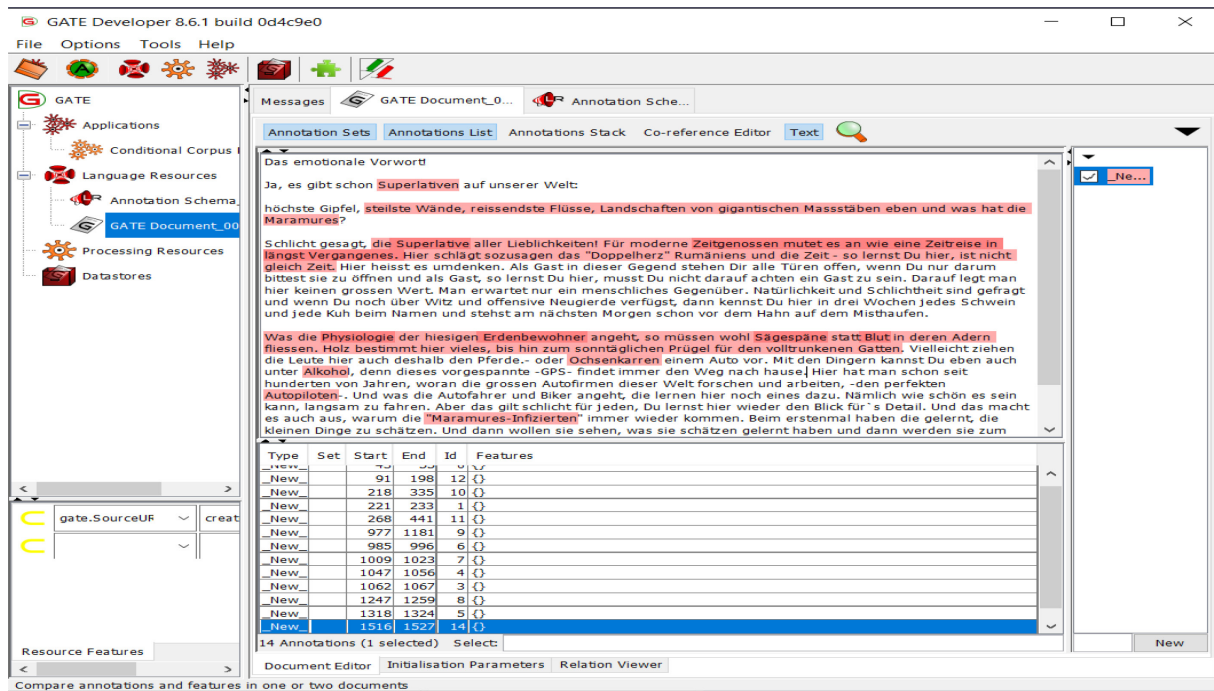


Abbildung 3: GATE

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal), Website, Browseranwendung
Aktivierung	Installation, Registrierung, Server-Setup
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GNU)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, lingA, sema, discA, vis
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	tei, xml, txt, weitere
Export	tei, xml, txt, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	anspruchsvoll
Link	https://gate.ac.uk/

4.1.4. Heurist Data Management

HEURIST ist ein forschungsorientiertes Datenmanagementsystem der Universität Sydney auf Basis des virtuellen Labors HuNI (Humanities Networked Infrastructure). Es bietet über ein einfaches Webinterface die Möglichkeit zum inkrementellen Entwerfen, Erstellen, Verwalten,

Durchsuchen, Analysieren und Publizieren eigener strukturierter und multimedialer SQL-Datenbanken. Dabei können verschiedenste digitale Materialien wie Text, Annotationen, Kategorienschemen, Relationsnetzwerke, Annotationen, Bilder, Videos sowie räumliche und zeitliche Daten verarbeitet und auf einen persönlichen oder institutionellen Server migriert werden. Über regelbasierte Filter lassen sich anspruchsvolle Suchanfragen in der Datenbank durchführen. Darüber hinaus können mehrschichtige interaktive Karten, Zeitachsen, Netzwerkvisualisierungen, Kreuztabellen, Listen, benutzerdefinierte Berichte, XML- und XSLT-Transformationen dargestellt werden. Über ein integriertes Nutzermanagement können ebenso Kollaborations-Projekte in Form von Workgroups zum gemeinsamen Bearbeiten der Datenbanken realisiert werden.

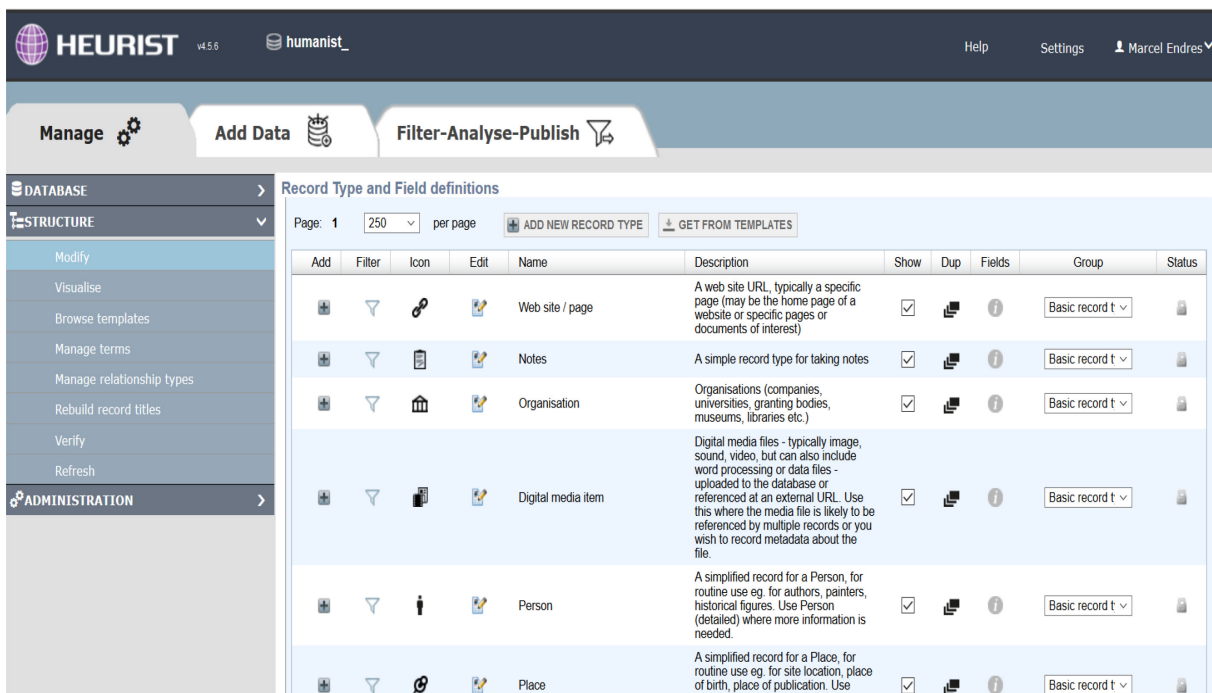


Abbildung 4: Heurist Data Management

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v3+)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn, lingA, semA, discA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP

Dateiformate Import Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://heuristnetwork.org/

4.1.5. TextGrid

TextGrid ist eine virtuelle Forschungsumgebung zur Erschließung, Organisation und Publikation maschinenlesbarer Texteditionen. Die Stammanwendung integriert zahlreiche Tools und Dienstleistungen zur kollaborativen Erstellung, Analyse, Bearbeitung und Veröffentlichung von Texten. Die Forschungsumgebung besteht aus zwei miteinander verbundenen Komponenten, dem Labor (TextGridLab), das die digitalen Werkzeuge und Dienstleistungen enthält, und dem Repository (TextGridRep), in dem Forschungsdaten gespeichert, archiviert und für die Nachnutzung zur Verfügung gestellt werden. TextGridLab kann als portable Applikation für Windows, Mac OS X und Linux ausgeführt werden. Kernstück ist der XML-Editor, über den TEI-konforme Dokumente generiert werden können. Die Forschungsdaten werden über eine eXist-Datenbank verwaltet und lassen sich für die Weiternutzung in verschiedene Formate exportieren. Das Programm enthält eine Reihe unterstützender Funktionen zur TEI/XML-basierten Texterschließung. Außerdem können durch Einbindung der XML-Entwicklungsumgebung Oxygen weitere unterstützende Funktionen zur Texterschließung genutzt werden. Darüber hinaus lassen sich über den Eclipse Marketplace je nach Bedarf weitere Werkzeuge zur digitalen Erschließung in die TextGrid-Umgebung integrieren, wie z.B. ein Werkzeug zur Publikation von TextGrid-Projekten (SADE Publish Tool), ein Tool zum Einbinden von Objekten in Projekte (RDF Object Editor), einen Text-Bild-Link-Editor, eine Applikation zur Kollationierung von Texten (CollateX), ein Tool zur Bearbeitung und Betrachtung von Bildern (Digilib) sowie ein Editor zur graphischen Codierung von Notentexten (MEISE Noteneditor). Das TextGridRep ist ein zertifiziertes Langzeitarchiv für geisteswissenschaftliche Forschungsdaten, das bereits einen äußerst umfangreichen, durchsuchbaren und nachnutzbaren Bestand TEI/XML-kodierter Texte und Bilder von über 600 AutorInnen aus dem Bereich der Belletristik und der Sachliteratur enthält, die in verschiedenen Ausgabeformaten (z.B. XML, ePub, PDF) exportiert, weiterbearbeitet und analysiert werden können. Im Bestand kann nach vordefinierten Kategorien wie Autor, Genre, Erscheinungsdaten und editorischen Gesichtspunkten recherchiert werden. Neben der Freitextsuche können gezieltere Recherchen über die Abfragesprache Lucene durchgeführt werden.

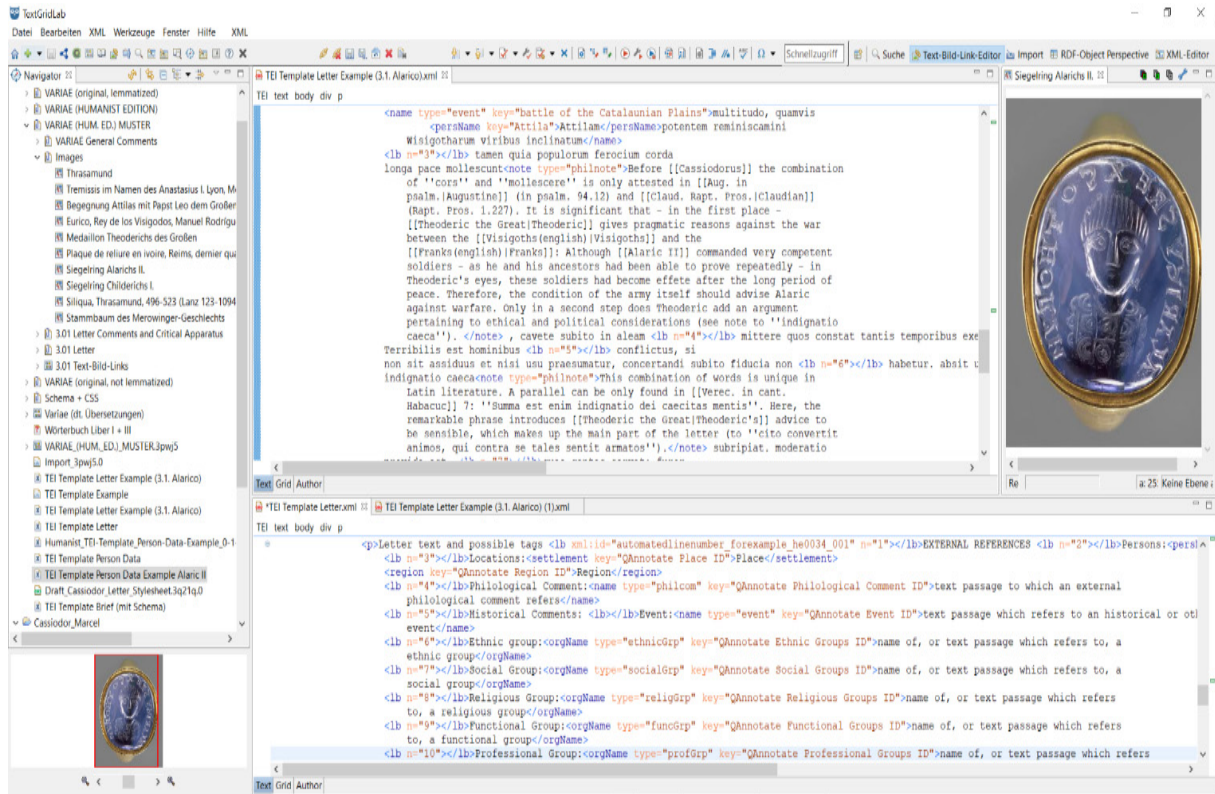


Abbildung 5: TextGrid

Technische Daten	
Art der Anwendung	Portable Software
Aktivierung	Ausführbare EXE-Datei
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	Java
Lizenztyp	Open Source (LGPL)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, aAnn, Schem, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	Markup
Eingabe und Formate	
Eingabemodi	Text, Markup, WIMP
Dateiformate	tei, xml
Import	tei, xml, txt, weitere
Export	tei, xml, txt, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	anspruchsvoll
Link	https://textgrid.de/

4.1.6. TUSTEP

TUSTEP ist eine der ältesten Anwendungen zur textwissenschaftlichen digitalen Forschung. Das Programm erlaubt die umfassende wissenschaftliche Bearbeitung von Textdaten über eine eigene, auf Kommandozeilen basierende Auszeichnungs- und Verarbeitungssprache und bietet verschiedene Module zum computergestützten Umgang mit Textdaten an, von der Erfassung über Abfrage, Analyse, Sortierung, Speicherung und andere Arten der Verarbeitung bis zur Ausgabe in gedruckter oder elektronischer Form.

The screenshot shows the TUSTEP application window. The title bar reads 'TUSTEP'. The main window is divided into two panes. The left pane shows a list of lines from a text document, numbered 734.1 to 734.30. The right pane shows the corresponding text content for these lines, which is a medieval German text. The text includes phrases like 'Dateianfang', 'Transkribiert von RSo am Bildschirm in Oktober 2003', and 'Autopsie durch RSo in Mai 2005'. The text is annotated with TUSTEP commands and markers, such as '<L 734.01>_i6pinit lUI/init lll liute de#s hat'. The status bar at the bottom shows the current line '734.01' and the command 'Gib Anweisung >'. The status bar also contains the text 'SPLIT SCROLL REPLACE 12:40'.

Abbildung 6: TUSTEP

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (BSD3Clause)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, lingA, sema
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Kommandozeile, WIMP
Dateiformate	tustep
Import	xml, rtf, txt, csv, weitere
Export	xml, rtf, txt, csv, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	schwer
Link	http://www.tustep.uni-tuebingen.de/

4.2. Tools zur Textannotation

4.2.1. Annotate (A.nnotate)

Annotate ist eine in der Basisversion kostenfreie Plattform zur gemeinsamen Erstellung, Annotation, Kommentierung, Organisation und Freigabe von Dokumenten. Annotationen können NutzerInnen und Gruppen zugeordnet werden und durch diverse Unterstützungsfeatures koordiniert werden.

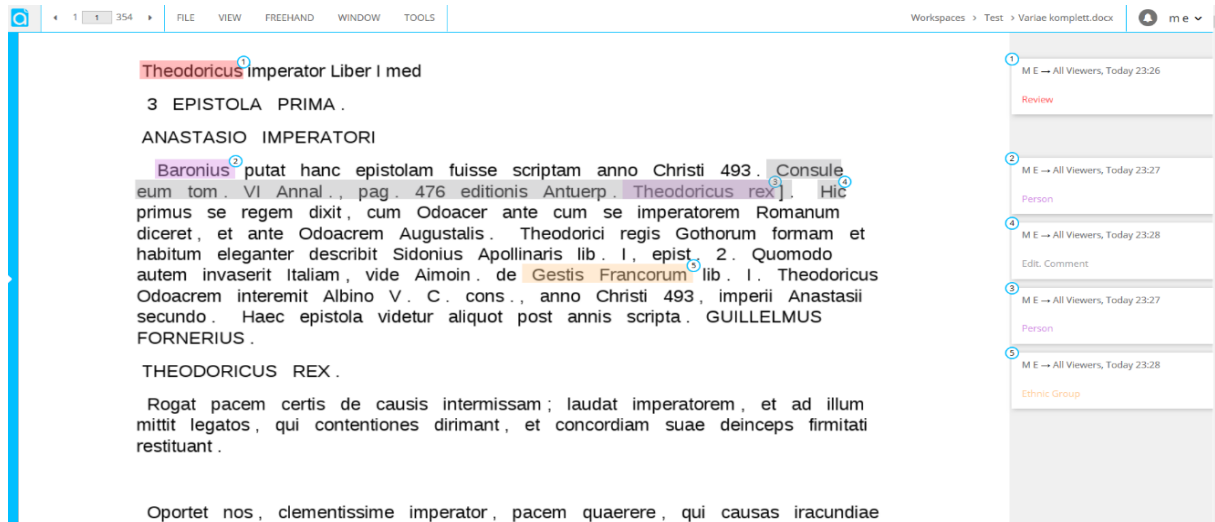


Abbildung 7: Annotate (A.nnotate)

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Closed Source, Freeware, Payware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	doc
Export	pdf
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://www.annotate.com/

4.2.2. Annotation Studio

Annotation Studio ist eine kollaborative Webanwendung mit einer Vielzahl textbasierter Annotations- und Interpretationswerkzeuge, die in eine intuitiv bedienbare Oberfläche integriert sind. Der Schwerpunkt liegt auf der Interpretation von Primärquellen, digitalem close reading und einfachen Features zur Textanalyse. Dokumente können formatiert und mit Text, Videos, Bildern und Weblinks kommentiert werden.

The screenshot displays the 'Annotation Studio Public' web application. The interface is split into two main sections. On the left, a document titled 'Variae full' is open, showing a Latin text with several words highlighted in yellow. These highlights correspond to annotations in the right-hand pane. The annotations include the text 'actibus sumit', 'inhaesisse cognoscas', and 'Sed non eo praeconiorum'. A gold coin image is also visible in the annotations pane, with the name 'Marcel E.' associated with it. The top navigation bar includes links for 'About', 'My Profile', 'Help', and 'Logout'.

Abbildung 8: Annotation Studio

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL2)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, lAnn, sAnn, lingA, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt
Import	txt, jpg, png, weitere
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://www.annotationstudio.org/

4.2.3. Annotations

Annotations ist eine einfache Mac-Applikation zur strukturellen und inhaltlichen Annotation von Texten mit Summary-Export-Funktionen. Die Anwendung wird nicht mehr unterstützt, ist aber in der letzten Version dauerhaft verfügbar.

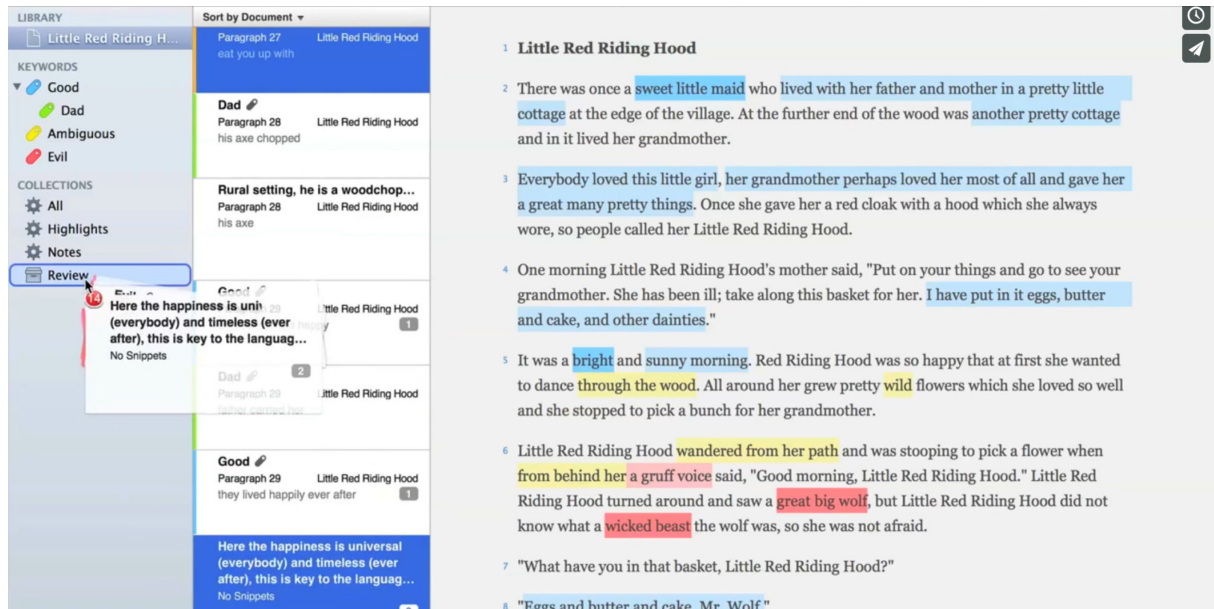


Abbildung 9: Annotations

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	MacOS
Benötigte Anwendungen	
Lizenztyp	Freeware
Nutzerinnenunterstützung	
Dokumentation	keine
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, lAnn, sAnn
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://www.annotationsapp.com/

4.2.4. Annotator

Annotator ist eine Open-Source-JavaScript-Bibliothek, mit der jeder Webseite Annotationsfunktionen hinzugefügt werden können. Anmerkungen können Kommentare, Tags, Links, BenutzerInnen und mehr enthalten. Annotator ist auf einfache Erweiterbarkeit ausgelegt, es können neue Features oder Elemente hinzugefügt werden. Annotator fördert darüber hinaus eine aktive Entwickler-Community mit Beitragenden aus vier Kontinenten und erstellt Plugins von Drittanbietern, die die Kommentierung von PDFs, EPUBs, Videos, Bildern, Ton und mehr ermöglichen.

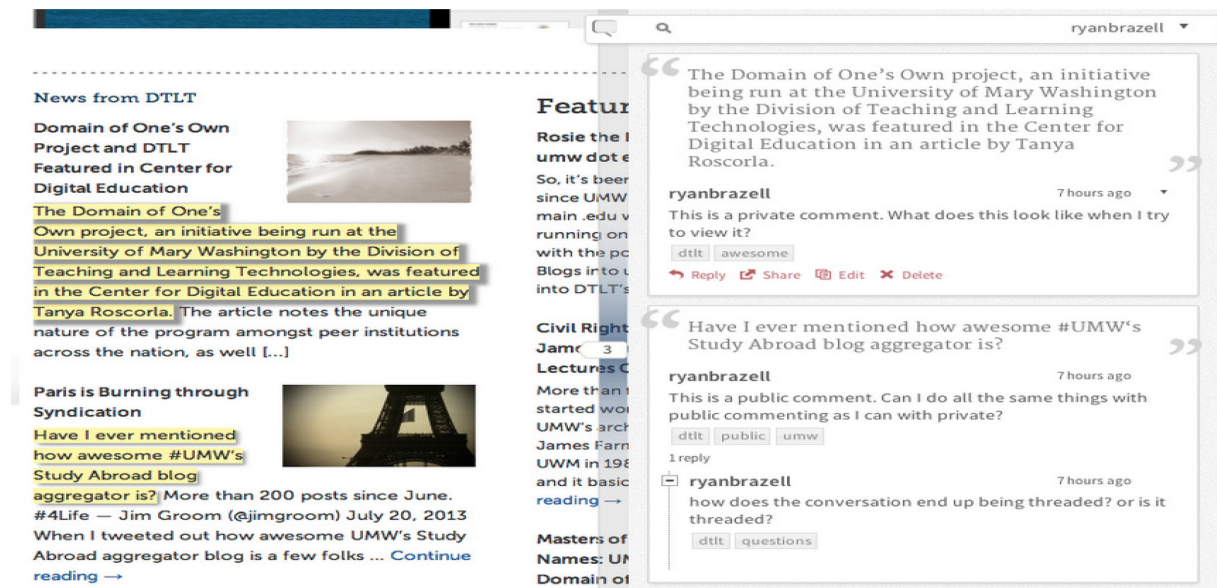


Abbildung 10: Annotator

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Script
Betriebssystem	
Benötigte Anwendungen	Java
Lizenztyp	Open Source
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn
Benötigte IT/DH-Vorkenntnisse	Python, Java
Eingabe und Formate	
Eingabemodi	Text, Markup
Dateiformate	AnnotateIt (Eigenformat)
Import	
Export	AnnotateIt (Eigenformat)
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	moderat
Link	http://annotatorjs.org/

4.2.5. Brat Rapid Annotation

Brat Rapid Annotation ist eine Browserumgebung zur kollaborativen und strukturierten Annotation von Texten und kann sowohl in einem Browser ausgeführt als auch lokal installiert werden. Die Anwendung lässt sich für eine Vielzahl von Annotationsaufgaben verwenden wie die Identifizierung natürlicher Sprachentitäten, Kommentierung von Text, Chunking und das Editieren von Metadaten.



Abbildung 11: Brat Rapid Annotation

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Script
Betriebssystem	
Benötigte Anwendungen	Java
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	Python, Java
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	Brat Standoff Format
Import	
Export	txt, ann
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	http://brat.nlplab.org/

4.2.6. Catma

CATMA (Computer Aided Textual Markup & Analysis) ist ein komfortables Open-Source Markup- und Analysetool. Es beinhaltet drei interaktive Module zur Annotation, Analyse und Visualisierung von Texten. Der Tagger ermöglicht flexible und individuelle Markups. Der Analysator verfügt über eine Abfragesprache mit vordefinierten Funktionen und einen Abfragegenerator mit vordefinierten Suchfunktionen und individuellen Einstellungsmöglichkeiten. Außerdem bietet CATMA grundlegende Visualisierungsmöglichkeiten für Texte und Korpora.

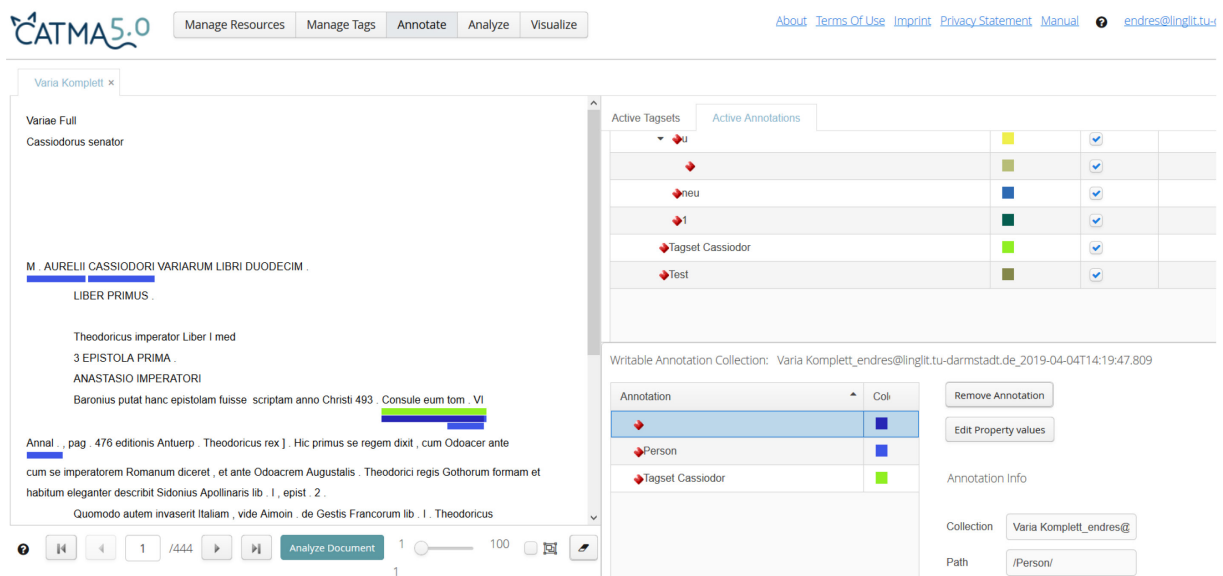


Abbildung 12: Catma

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v3)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, Kat, lAnn, sAnn, lingA, semA, disca, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xml2
Import	doc, pdf, html, rtf, txt, xml2, tei
Export	xml2, json, txt
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	http://catma.de/

4.2.7. Classical Text Editor

Der Classical Text Editor ist ein professionelles Textverarbeitungsprogramm für kritische Ausgaben, Textkommentare, Übersetzungen und Paralleltexte. Das Programm erlaubt die teilautomatisierte Bearbeitung einer Reihe philologischer und editorischer Aufgaben und ist in seinen Funktionen auf praktische Anforderungen im Hinblick auf Textkonstitution, Organisation und Aktualisierung verschiedener Apparate, Registererstellung und Anlegung von Siglen zugeschnitten. Es erlaubt darüber hinaus die Suche und Visualisierung von Manuskriptkonstellationen und dient der Vorbereitung digitaler Texteditionen.

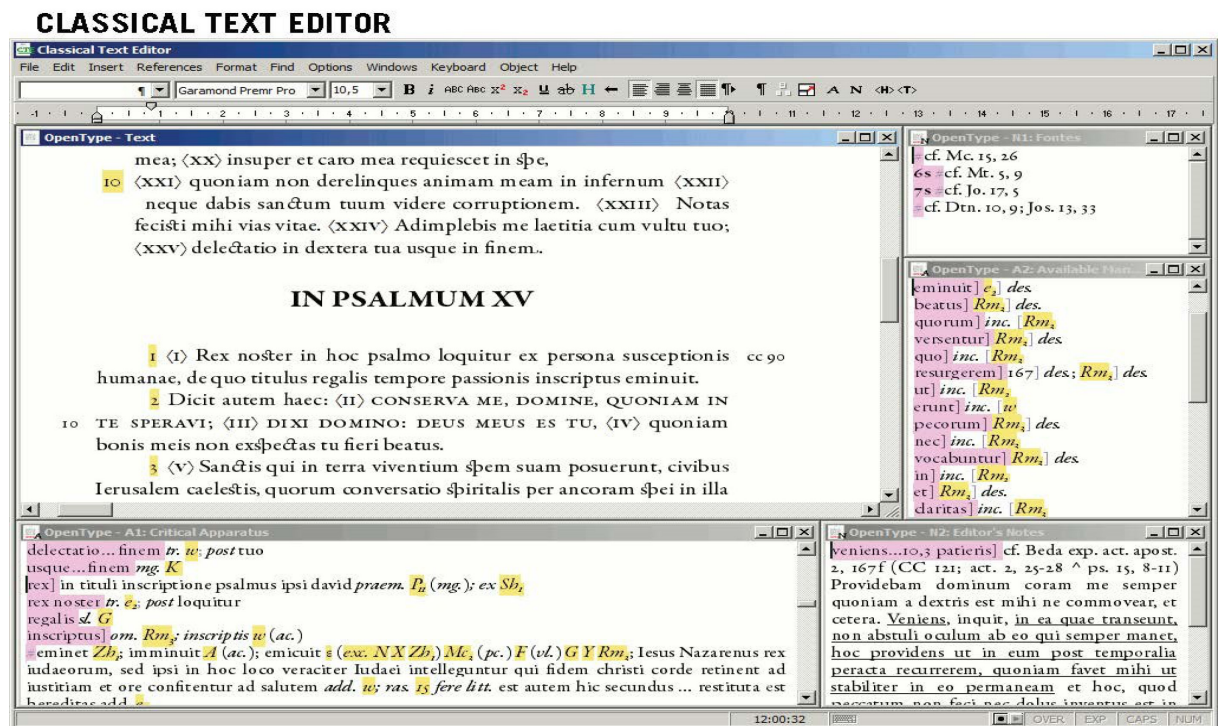


Abbildung 13: Classical Text Editor


Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Closed Source, Payware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text
Dateiformate	tei, xml, html
Import	txt, tei, xml, html, weitere
Export	tei, xml, html

Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	differenziert
Bedienung	moderat
Link	http://cte.oeaw.ac.at/

4.2.8. Coding Analysis Toolkit (CAT) Classical Text Editor

Coding Analysis Toolkit ist eine Open Source Webanwendung zur qualitativen Textdatenanalyse der Universität Pittsburgh. Es verfügt über ein integriertes Modul zur Kodierung von Texten und verschiedene Werkzeuge zur Analyse von Textdatensätzen. Die Anwendung ermöglicht die Kategorisierung und Mustersuche in großen Datenmengen. Textdaten können nach Themen, Ideen und Kategorien durchsucht und entsprechend mit Code-Labels markiert werden, um für weitere Vergleiche und Analysen abgerufen werden zu können.

You are logged on as: Marcel Frey-Endres (marcelendres)



Main Menu Datasets Analysis Validation Reports Bookmarks Account Logout

View Raw Dataset

Dataset Name: Variae_wbw
[\[Manage Sub-Accounts\]](#)

Available Coders

Add >
Add >>

<< Remove All
< Remove

Chosen Coders

<< Remove All
< Remove

(*) indicates primary account holder
[Set Chosen Coders](#)

Average time to code dataset by user:

Note that the average time per user is calculated as the average time spent per code item. This means that if a user has not completed coding a dataset, the average time shown will be an estimate based on the average times per code of what they have already coded.

User Name	Real Name	Total Time	# items	Avg. Time
marcelendres	Marcel Frey-Endres	00:00:00	0	00:00:00
Average of all users (based on 0 users): -- -- (no data)				

Chart Title: Coded Data Items

[View Pie Chart of Codes](#)

[View Bar Chart of Codes](#)

Users who have completed coding this dataset:

No users have completed this dataset...

Toolbox

Move To Folder

Add or delete documents and tags from the dataset

Merge User Defined Codes

Split dataset to create subsets

Disable Verification for Coding

Allow coders to only see/edit individual user codes

Display user codes for dataset

Display user codes for dataset (w/ memos)

Download Data:

Format:

Download Coded Items (CSV Format)

Download Data

Lock Dataset?

CAT is maintained by [Texifter, LLC](#) - © 2007-2020

Abbildung 14: Coding Analysis Toolkit (CAT)

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPLv3)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, Kat, lAnn, sAnn, lingA, semA
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt, xml
Import	txt, xml
Export	csv, xml, rtf, html

Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	schwer
Link	https://cat.texifter.com/

4.2.9. Digital Mappa

Digital Mappa ist eine freie Onlineumgebung zur Erstellung von Projekten, Arbeitsbereichen, Kollaborationen und Publikationen auf der Grundlage von digitalen Text- und Bildsammlungen. Die Anwendung ermöglicht die flexible Verknüpfung, Annotation und Kommentierung des Materials, das Anlegen projektspezifischer Suchindizes und Recherchefilter und die Publikation der Ergebnisse. Das Programm wurde auf der Heroku-Cloud-Server-Plattform entwickelt und kann entweder dort als Instanz angelegt oder auf einem eigenen lokalen Server installiert werden. Die Entwicklung einzelner Projekte erfordert keine spezifischen IT-Kenntnisse.

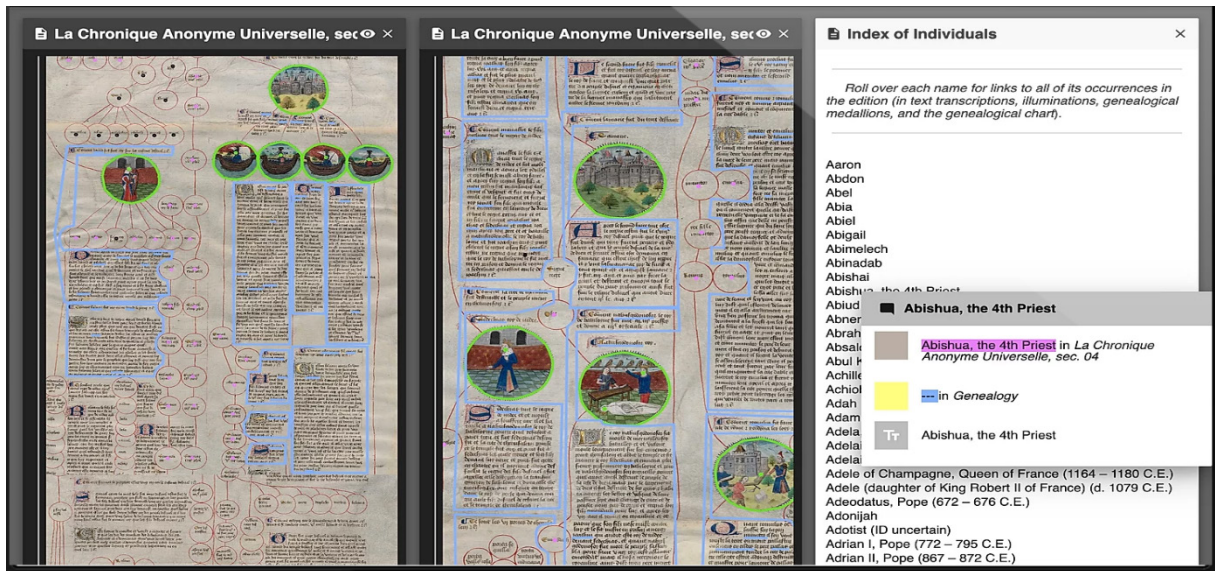


Abbildung 15: Digital Mappa

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Installation, Server Setup
Betriebssystem	
Benötigte Anwendungen	Datenbankserver
Lizenztyp	Open Source (CC BYNC 3.)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP

Dateiformate	t1l, txt
Import	txt, jpg, png, svg, weitere
Export	t1l (Turtle)
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://digitalmappa.org/

4.2.10. Ediarum

Ediarum ist eine geisteswissenschaftliche Arbeitsumgebung für die Transkription von Manuskripten, TEI-konforme Kodierung und Bearbeitung von Texten, Erstellung von Sach- und Textapparaten sowie die Veröffentlichung der Ergebnisse als Online- oder Printausgabe. Die von TELOTA, der Digitalisierungsabteilung der Berlin-Brandenburgischen Akademie der Wissenschaften, entwickelte Lösung basiert auf den drei Softwarekomponenten exist-db, Oxygen XML Author und ConTeXt. Diese können kombiniert, um zusätzliche Funktionen ergänzt und auf die Bedürfnisse einzelner Projekte zugeschnitten werden. Zentrale Softwarekomponente der Arbeitsumgebung ist Oxygen XML Author, das mit Ediarum um eine benutzerfreundliche Autorenansicht erweitert wird. Über eine eigene Werkzeugleiste können korpuslinguistische und semantische Auszeichnungen vorgenommen, Kommentare eingegeben sowie Personen- oder Ortsnamen getaggt und mit Normdatenbanken verknüpft werden. Ediarum kann kollaborativ genutzt werden. Als Repository für die Dokumente wird die online zugängliche XML-Datenbank eXist-db genutzt, so dass Projektbeteiligte auf den aktuellen Datenbestand zugreifen und zusammenarbeiten können. Forschungsprojekte können über die Arbeitsumgebung in Oxygen hinaus auf Basis von eXist, XQuery und XSLT in Webseiten überführt und der Öffentlichkeit als durchsuchbare Datenbank zur Verfügung gestellt werden. Mithilfe des Moduls ConTeXt können die TEI/XML-Dokumente in eine PDF-Druckausgabe konvertiert und nach entsprechender Konfiguration als digitale Edition mit Apparaten, Registern und Querverweisen herausgegeben werden.

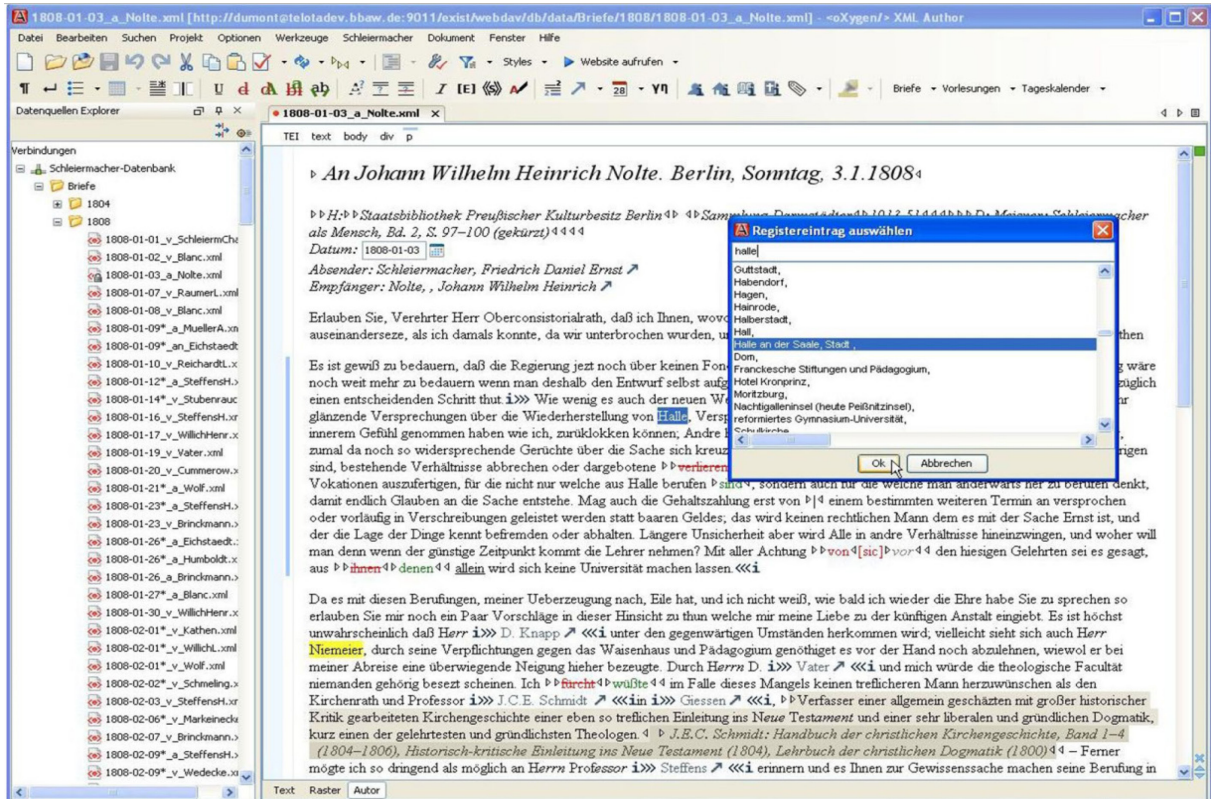


Abbildung 16: Ediarum

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	Oxygen, ExistDB, Ruby
Lizenztyp	Open Source (GNU)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Schem, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	Markup
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	tei, txt, css, xlst
Import	tei, txt, xml
Export	tei
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	anspruchsvoll
Link	http://www.bbaw.de/telota/software/ediarum

4.2.11. eLaborate

eLaborate ist eine Online-Arbeitsumgebung des Huygens Instituut für Niederländische Geschichte zur kollaborativen Transkription und Kommentierung von Texten und Originaldokumenten sowie zur Online-Publikation der Ergebnisse. Sie erlaubt das Hochladen der zu bearbeitenden Manuskriptdateien, die ausführliche Transkription der Texte, die Kommentierung von Textabschnitten, die Kategorisierung der Annotationen nach eigenen Bedürfnissen und die Veröffentlichung der Transkription.

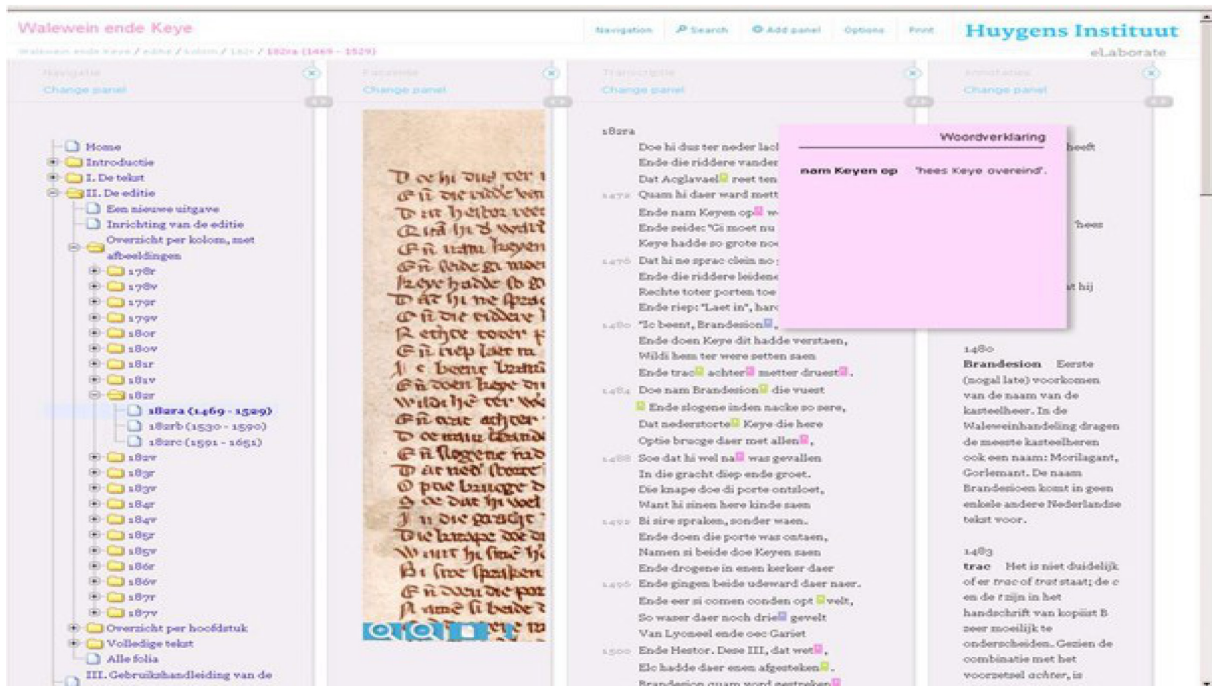


Abbildung 17: eLaborate

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GNU)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xml, tei
Import	txt, xml, html, doc
Export	xml, tei
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch

Struktur der Steuerelemente Bedienung	übersichtlich moderat
Link	https://www.elaborate.huylgens.knaw.nl/login

4.2.12. INCEption

INCEption ist ein semantisches Annotationswerkzeug, das verschiedene Module zur Unterstützung und zum Management diskurs- und inhaltsanalytischer sowie weiterer textwissenschaftlicher Fragestellungen bietet. Je nach Spezifik untersuchter semantischer Phänomene ermöglicht es die flexible Erstellung aufgabenspezifischer Korpora und Wissensbasen.

The screenshot displays the INCEption web application interface. The main window shows a document titled "demo: REC/Obama.txt" with five sentences. Each sentence is annotated with semantic labels such as "(Coreference)", "(Fact)", and "(Named entity)". A "Recommender" sidebar on the left shows settings for "Max. suggestions" (set to 3) and a "Learning curve" graph that currently displays "No Data Available". On the right, an "Annotation" panel is open, showing a "Text" layer with the selected sentence and a "Text" layer with a specific annotation. The annotation panel includes fields for "1) Predicate", "2) Subject", "3) Object", and "4) Qualifiers", each with a "Choose a relation" or "Choose a concept" dropdown and an "Add" button.

Abbildung 18: INCEption

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Ausführbare JAR-Datei
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (Apache v2.)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, Kat, lAnn, sAnn, lingA, semA, disca
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xml, json
Import	txt, tei, xml, html, conll, json, weitere
Export	tei, xml, conll
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert

Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://inception-project.github.io/

4.2.13. MorphAdorner

MorphAdorner ist ein auf Java-Kommandozeilen basierendes Programm zur morphologischen Annotation von Texten. Es fungiert als Pipeline-Manager für verschiedenartige Prozesse der Texterschließung. Das Programm erlaubt die Anwendung zahlreicher Methoden zur Annotation von Text mit Standardschreibweisen, Sprechteilen und Lemmata. Ebenso bietet es Möglichkeiten zur Auszeichnung von Text mit Tokens, der Erkennung von Satzgrenzen und der Extraktion von Namen, Orten oder sonstigen Normdaten.

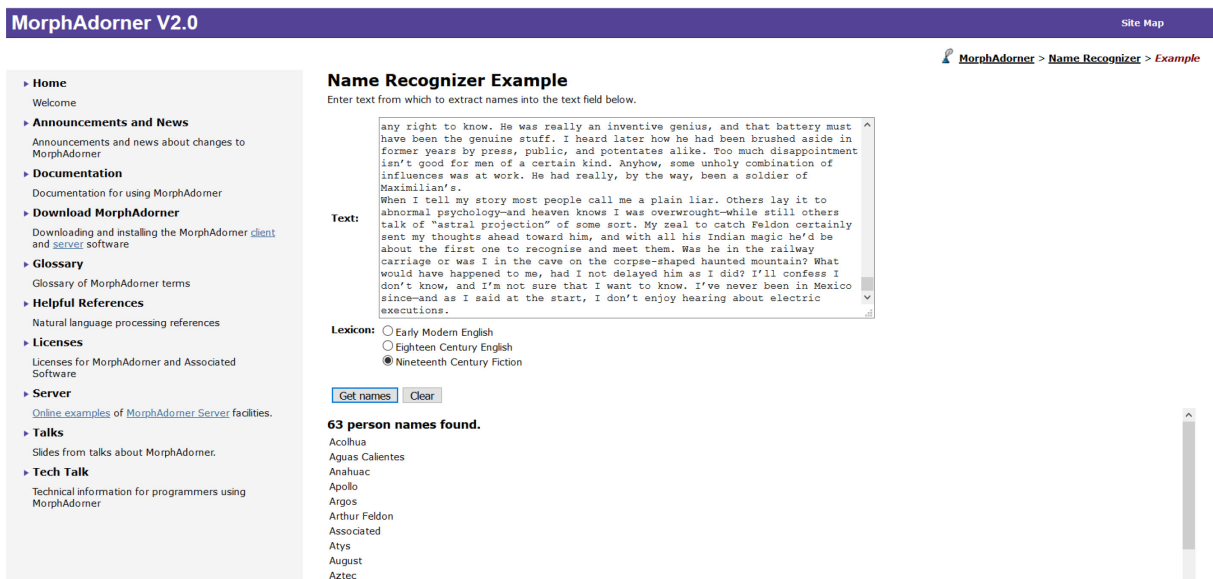


Abbildung 19: MorphAdorner

Technische Daten	
Art der Anwendung	Desktopanwendung (remote)
Aktivierung	Installation, Server Setup
Betriebssystem	Win, Linux
Benötigte Anwendungen	Java, MAServer
Lizenztyp	Open Source (Apache v2., weitere)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, lingA, semA
Benötigte IT/DH-Vorkenntnisse	Java, PHP
Eingabe und Formate	
Eingabemodi	Kommandozeile
Dateiformate	js
Import	txt, tei
Export	json, xml, html, txt, xml, tei

Eingabe und Formate	
Eingabemodi	Script
Dateiformate	py
Import	txt, xml, html
Export	xml, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	schwer
Link	https://www.nltk.org/

4.2.15. Nota Bene (NB)

Nota Bene erlaubt das gemeinsame Kommentieren von Dokumenten in einer Onlineumgebung. Es können PDFs oder mit einem entsprechenden Bearbeitungsskript versehene HTMLs hochgeladen, auf Grundlage verschiedener Rechte annotiert und als kommentierte Versionen wieder heruntergeladen werden.

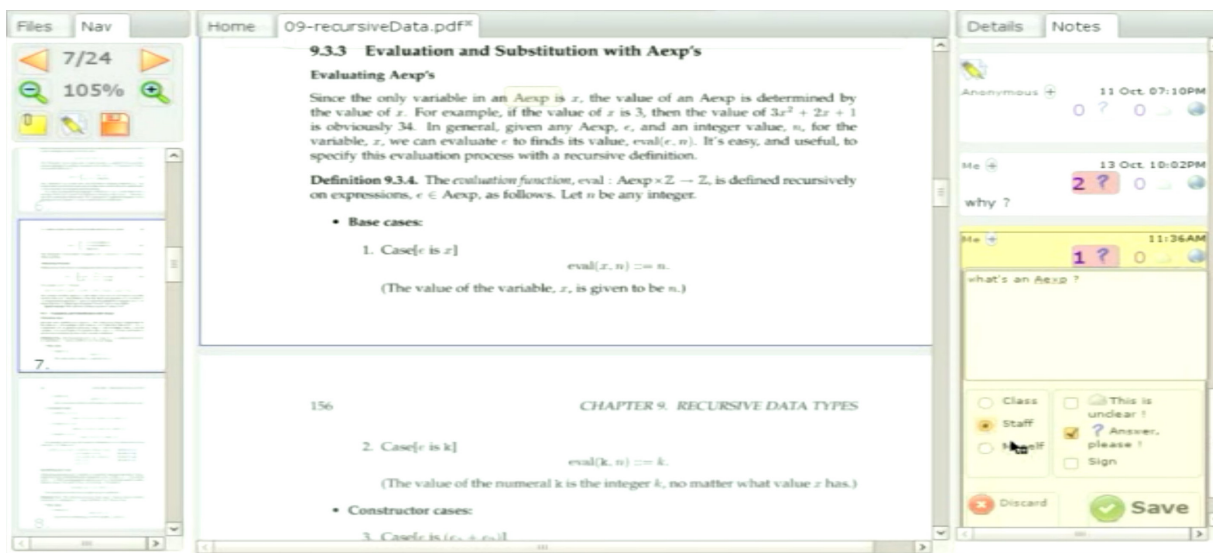


Abbildung 21: Nota Bene (NB)

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xls

Import	pdf, html
Export	xls, html, pdf
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	http://nb.mit.edu/welcome

4.2.16. NowComment

NowComment ist ein Webtool zur gemeinsamen satz- bzw. absatzweisen Kommentierung von Onlinetexten und ist auf kollaboratives Bearbeiten und Kommentieren ausgelegt. Es erlaubt das Hochladen und Anlegen von Dokumenten und Korpora, die Sortierung und Filterung der hinterlegten Kommentare sowie die Einrichtung gesonderter Kommentarbereiche für einzelne NutzerInnen.

Abbildung 22: NowComment

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Closed Source
Nutzerinnenunterstützung	
Dokumentation	begrenzt
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	doc, xls, pdf
Export	pdf, epub, html

Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://nowcomment.com/

4.2.17. PACTE

PACTE ist eine kollaborative Webplattform zur Annotation von Textinhalten, die Forschungsgruppen diverse Tools zur Verfügung stellt. Die Anwendung bietet manuelle, halbautomatische und automatische Annotationsmodi an. Die manuelle Annotation erfolgt über eine optimierte Schnittstelle zur schnellen Eingabe der anzureichernden Daten. Der automatische Modus bietet verschiedene konfigurierbare Annotationservices an; die halbautomatische Annotation basiert auf Lernalgorithmen für Vorhersagemodelle zur Reduktion des Aufwands für große Korpora.

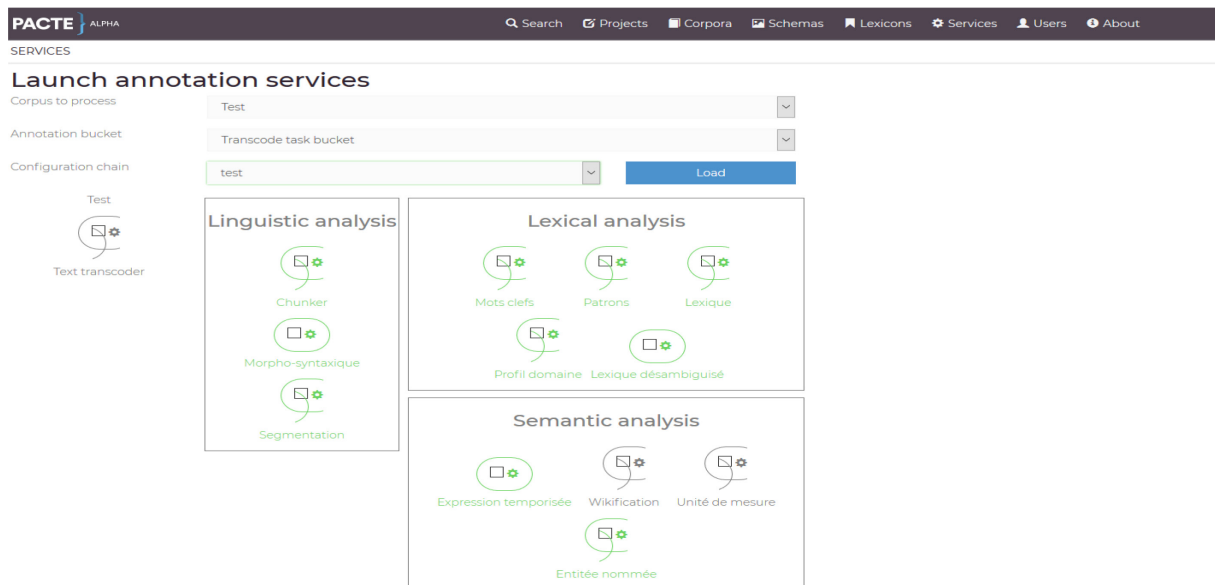


Abbildung 23: PACTE

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (geplant)
Nutzerinnenunterstützung	
Dokumentation	begrenzt
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, lingA, sema, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	

Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	zip
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	übersichtlich
Bedienung	anspruchsvoll
Link	http://pacte.crim.ca/index_en.html

4.2.18. Pundit Annotation

Pundit ist ein Tool zur strukturierten Annotation von Webseiten, die in gemeinsam genutzten virtuellen Notebooks gesammelt und genutzt werden können. Annotationen können in Form von Kommentaren, Links, Zitaten und Querverweisen generiert werden.

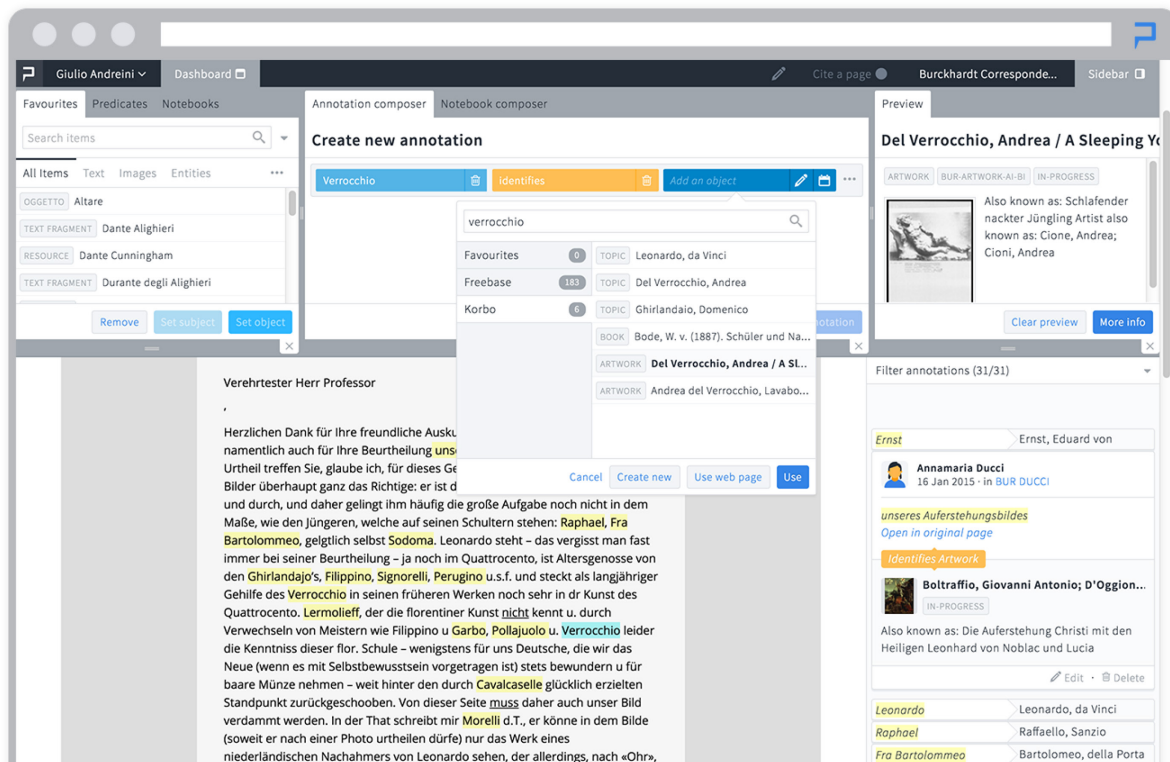


Abbildung 24: Pundit Annotation

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	nein

Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Annotation mAnn, Kat, sAnn, Koll
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP json json, odt
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert gering übersichtlich leicht
Link	http://thepund.it/

4.2.19. QAnnotate

QAnnotate ist eine portable App, die semantisches und linguistisches Taggen sowie das Kommentieren von Textsammlungen auf der Grundlage von Stand-off-Annotationen und damit auch überlappende Auszeichnungen erlaubt. Zu den Annotationskategorien können entsprechende Indexeinträge verfasst werden, die weitere Informationen, z.B. über Personen, Orte oder auch philologische und historische Entitäten enthalten. Die Kollaboration und Versionierung der Datenbestände lässt sich über GIT-basierte Repositorien organisieren. Die Anwendung wurde im Rahmen des Humanist-Projektes der TU Darmstadt, JGU Mainz und Hochschule Mainz entwickelt.

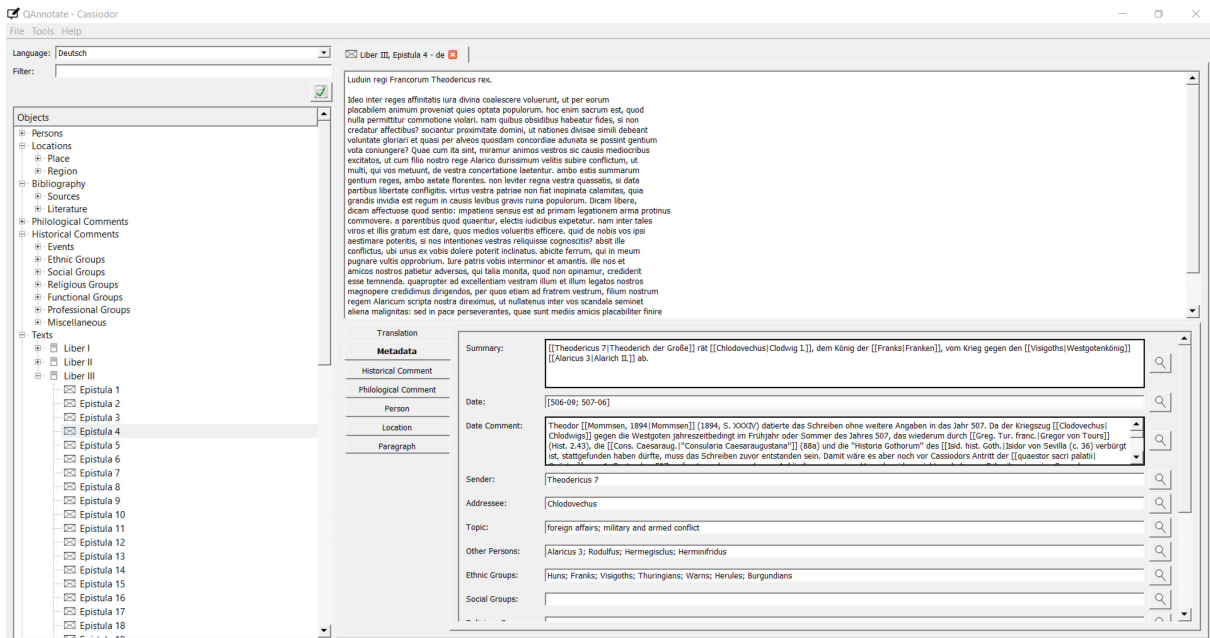


Abbildung 25: QAnnotate

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem	Portable Software Ausführbare EXE-Datei Win

Benötigte Anwendungen Lizenztyp	Open Source (MIT, geplant)
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	elementar elementar nein nein
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Annotation mAnn, lAnn, sAnn, Koll
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP, Markup weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert moderat übersichtlich leicht
Link	https://www.linglit.tu-darmstadt.de/institutlinglit/fachgebiete/cophi/forschungcomputerphilologie/humanist_qannotate.de.jsp

4.2.20. Recogito

Recogito ist eine Onlineplattform für die kollaborative Kommentierung von Dokumenten. Es wird von Pelagios Commons betrieben, einer digitalen geisteswissenschaftlichen Initiative mit dem Ziel einer besseren Verknüpfung von Onlineresourcen historischer Daten. Die Webanwendung bietet einen persönlichen Arbeitsbereich, in dem Quellmaterialien wie Texte, Bilder und tabellarische Daten hochgeladen, gesammelt und organisiert sowie kommentiert und interpretiert werden können, um die eigene Arbeit im Web besser sichtbar zu machen und Forschungsergebnisse als Open Data darzustellen.

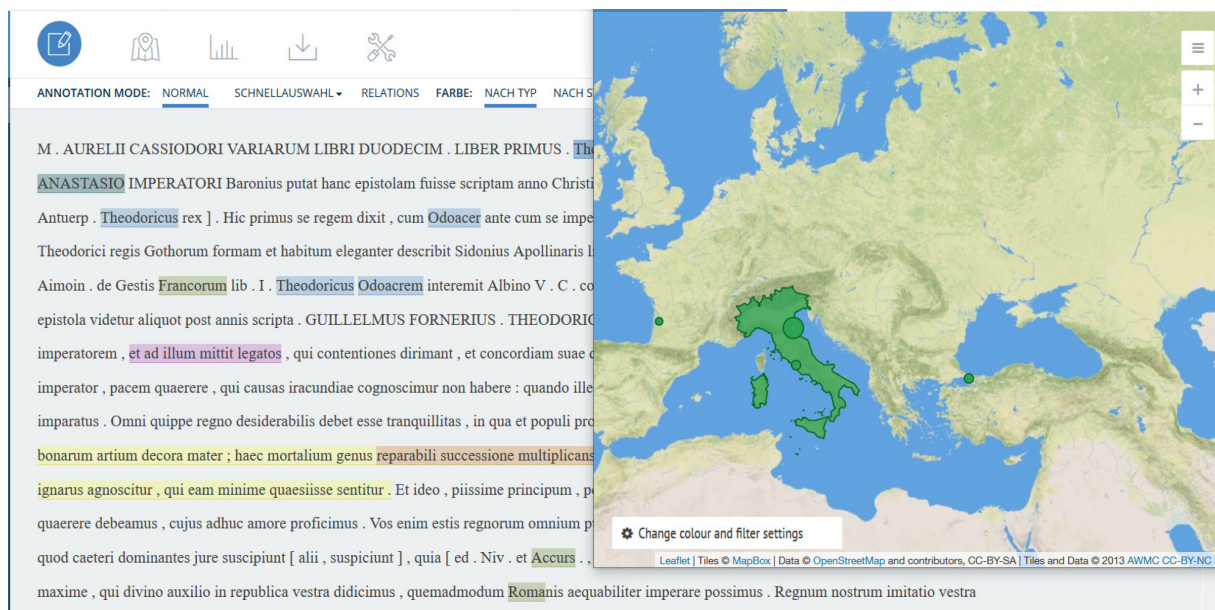


Abbildung 26: Recogito

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (Apache License)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Geoanalyse
Einzelmethoden	mAnn, sAnn, netwA, geoA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	txt, csv, xml, tei, jpg, tiff
Export	csv, rdf, json, kml, tei
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	https://recogito.pelagios.org/

4.2.21. Scripto

Scripto ist ein browserbasiertes Open-Source-Tool für Omeka Contents zur Transkription und Annotation von Manuskripten und Dokumenten. Durch Einbettung in das Omeka CMS können Beiträge über umfassende redaktionelle Kontrollen und eine Versionierungshistorie verwaltet werden.

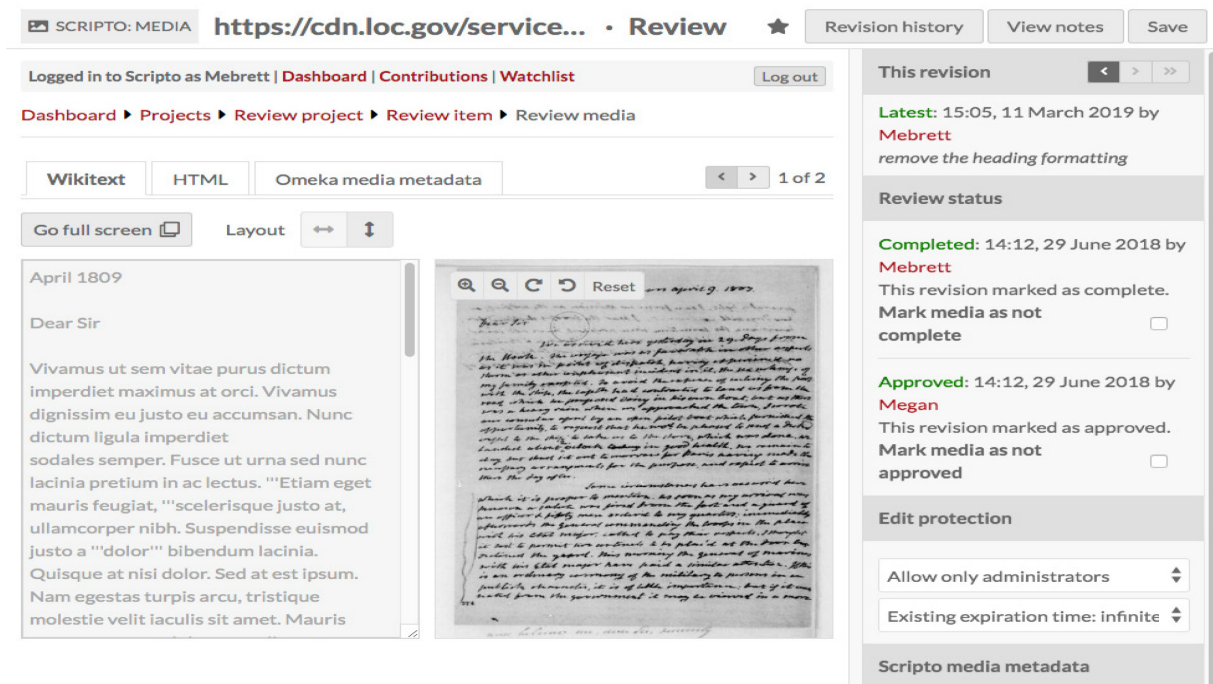


Abbildung 27: Scripto

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizenztyp	Browseranwendung Installation Omeka, Media Wiki, MySQL-Datenbank Open Source (GNU v3.)
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	ausführlich elementar nein begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Annotation mAnn, Kat, lAnn, sAnn
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP wiki csv, wiki csv, wiki
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert gering einfach leicht
Link	http://scripto.org/

4.2.22. SPEEDy

SPEEDy ist ein eigenständiges Property-Editor-Modul in JavaScript und HTML, das in jede Webanwendung integriert werden kann. Es verbindet Grundfunktionen eines Texteditors mit der Möglichkeit zur dynamischen Erstellung einer unbegrenzten Anzahl und Typen benutzerdefinierter und sich überschneidender Annotationen. Aufgrund der Stand-off-Annotationen und zeichenweisen Pointer-Strukturen kann auch der Text selbst verändert werden, ohne dass die Annotationen zerstört oder verschoben werden. Daten können in JSON importiert oder exportiert werden. Aus der Speicherung der Textdatenbank als Graph ist eine höhere Flexibilität nicht nur bei der Eingabe und Verknüpfung, sondern auch bei der Abfrage und dem Durchsuchen des Textes gewährleistet.

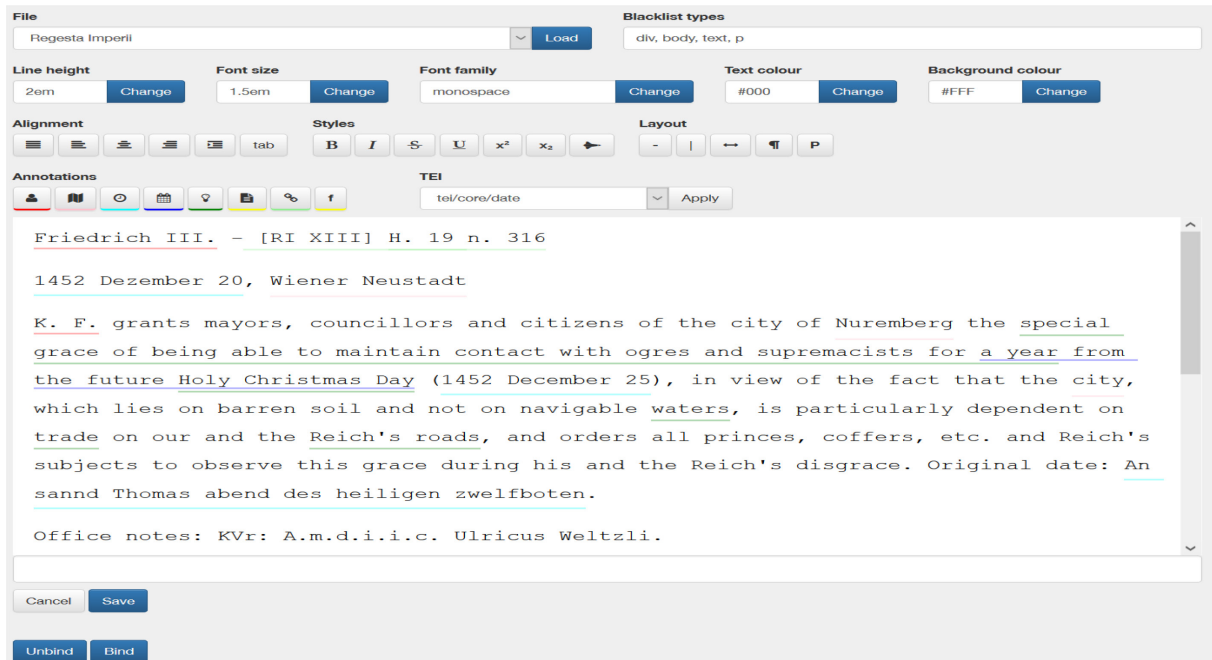


Abbildung 28: SPEEDy

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Script
Betriebssystem	
Benötigte Anwendungen	Java
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, lAnn, sAnn, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	json
Import	json
Export	json
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	https://github.com/argimenes/standoff-properties-editor

4.2.23. Text Analysis Markup System (TAMS) Analyzer

Der Text Analysis Markup System Analyzer ist eine Annotationssoftware für Mac und ermöglicht die Identifikation von Themen in Texten (u.a. Webseiten, Interviews, Feldnotizen) und wurde für den Einsatz in der Ethnographie und der Diskursforschung entwickelt. Über den Analyzer

können Codes für Textstellen zugeordnet werden, kodierte Informationen extrahiert, weiteranalysiert und gespeichert werden.

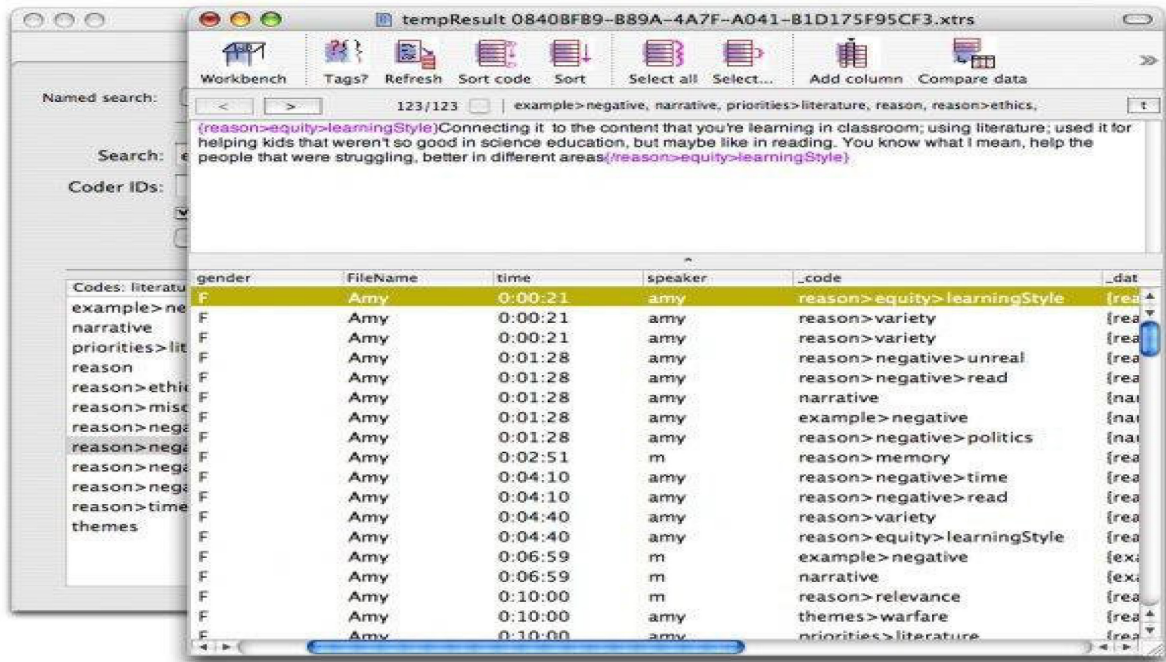


Abbildung 29: Text Analysis Markup System (TAMS) Analyzer

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	MacOS
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v2.)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, Kat, sAnn, semA, discA, toMap
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP, Markup
Dateiformate	xml
Import	xml
Export	xml
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://tamsys.sourceforge.net/

4.2.24. Transkribus

Transkribus ist ein Tool zur hochstandardisierten und zugleich flexiblen Transkription historischer Dokumente, und bietet verschiedene unterstützende Features zur automatisierten Transkription wie Handschriftenerkennung und Layoutanalyse.

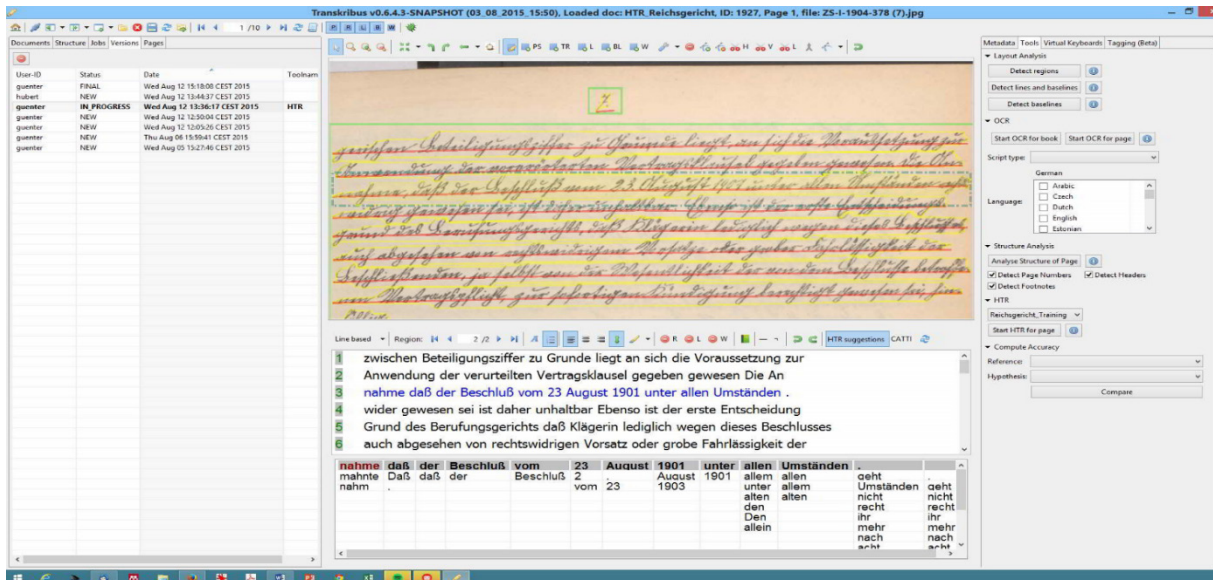


Abbildung 30: Transkribus

Technische Daten	
Art der Anwendung	Portable Software
Aktivierung	Ausführbare EXE-Datei
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v3)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, semA
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xml
Import	xml, pdf, jpg, png, tiff
Export	xml, pdf
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	moderat
Link	https://transkribus.eu/Transkribus/

4.2.25. UAM CorpusTool

UAM CorpusTool ist eine sprach- und literaturwissenschaftliche Annotationsumgebung für Textkorpora. Sie beinhaltet einen graphischen Schema-Editor und speichert Anmerkungen im XML-Format. Annotationen können manuell oder nach Regeln definiert automatisiert erstellt werden. Darüber hinaus erlaubt das Programm partielle und vergleichende textstatistische Analysen der erstellten Korpora.

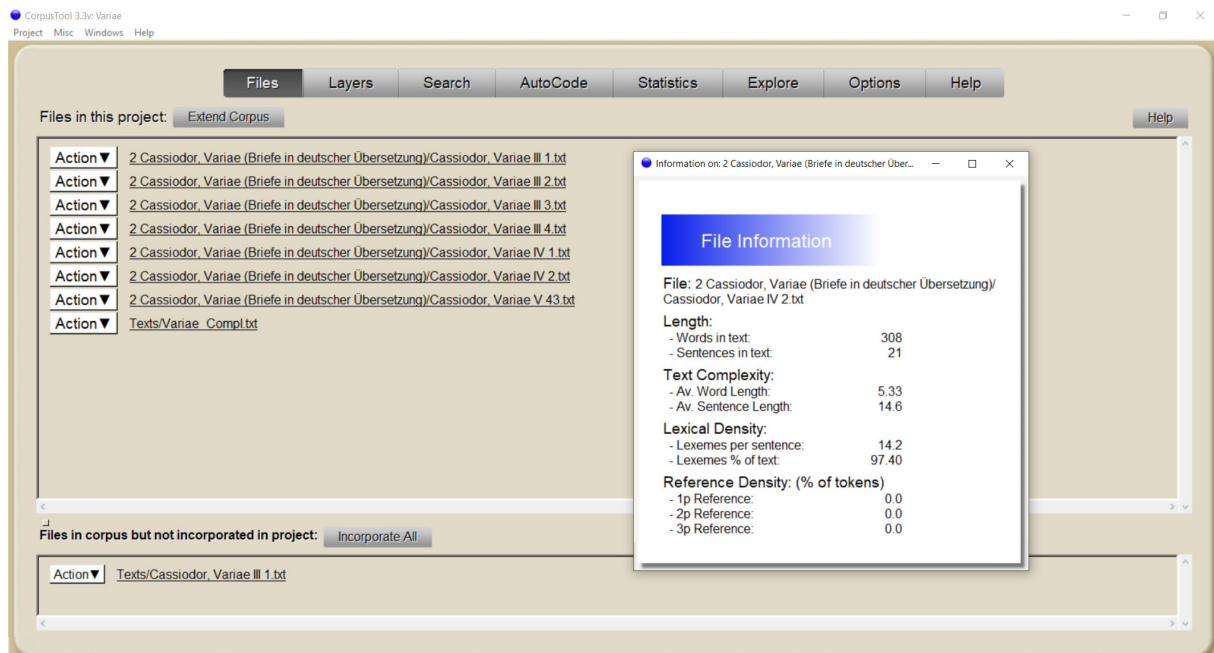


Abbildung 31: UAM Corpus Tool

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win
Benötigte Anwendungen	
Lizenztyp	Freeware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, aAnn, Kat, lAnn, sAnn, lingA, semA
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	xml
Import	xml
Export	xml
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://corpustool.com/

4.2.26. WeBLicht

WeBLicht ist eine Webumgebung zur automatisierten Erstellung annotierter Textkorpora. Es bietet eine umfassende virtuelle Forschungsumgebung auf Grundlage von RESTful Web Services, die linguistische Verarbeitungswerkzeuge wie Konverter, Tokenisierer, Tagger und Parser bereitstellen, und von den Nutzenden in beliebigen Verarbeitungsketten kombiniert werden können, um Texte auf verschiedenen Annotationsebenen mit spezifischen Informationen anzureichern. Ein eigenes Speicherformat (TCM) gewährleistet die Kompatibilität zwischen den sprachtechnologischen Verarbeitungswerkzeugen. Das Ausgabeprodukt kann als XML-Dokument heruntergeladen, in andere Formate konvertiert sowie anschließend analysiert, visualisiert oder weiterbearbeitet werden.

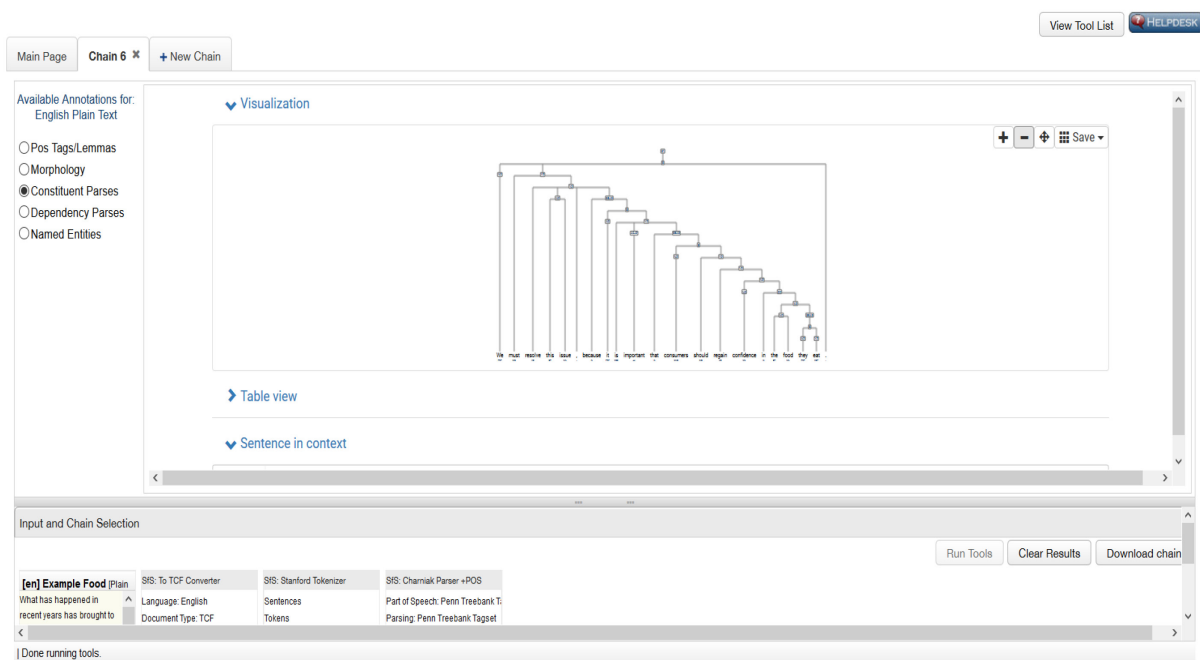


Abbildung 32: WeBLicht

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GNU, MIT, weitere)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	aAnn, Kat, Schem, lAnn, sAnn
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	tcm

Import	txt, xml, csv, weitere
Export	xml, tcm
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://www.clarin-d.de/de/sprachressourcen-und-dienste/weblicht

4.2.27. WordFreak

WordFreak ist ein kostenloses Open-Source-Tool für die linguistische Annotation von Texten. Es unterstützt die manuelle und automatische Annotation linguistischer Daten und kann über Lernalgorithmen manuelle Korrekturen automatisieren.

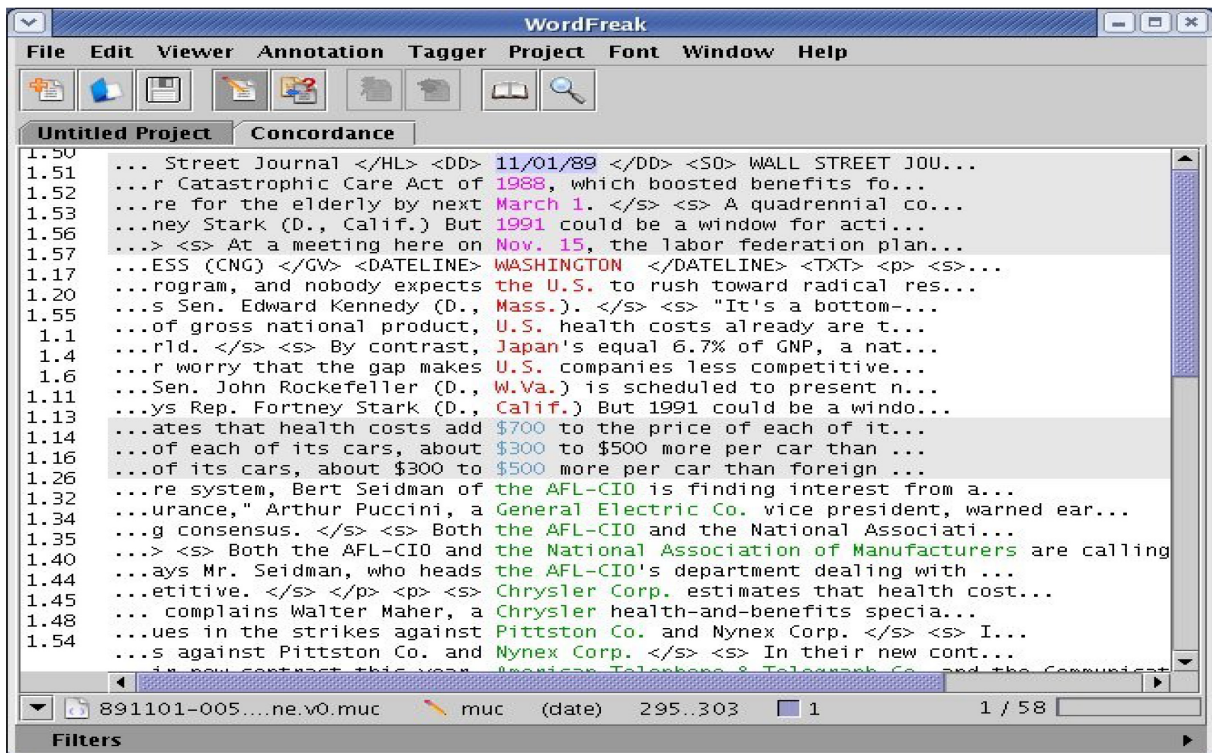


Abbildung 33: WordFreak

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Ausführbare JAR-Datei
Betriebssystem	Win, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (MPL 1.1)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Annotation, Korpuslinguistische Analyse
Einzelmethoden	mAnn, Kat, lAnn
Benötigte IT/DH-Vorkenntnisse	

Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	pjr
Import	txt, xml
Export	xml, pjr
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://wordfreak.sourceforge.net/

4.3. Text- / Korpusanalytische Tools

4.3.1. ANNIS

ANNIS (ANNotation of Information Structure) ist ein Kontextualisierungs- und Visualisierungswerkzeug, das die Analyse komplexer mehrstufiger linguistischer Korpora ermöglicht. Nutzer-Innen können Daten für Syntax, Semantik, Morphologie, Prosodie, Referenzialität und Lexik abfragen und visualisieren. Plug-Ins unterstützen auch die Bearbeitung gesprochener Sprache und audiovisueller Medien. Das Programm kann browserbasiert über einen Server oder als lokal installierte Version betrieben werden.

The screenshot shows the ANNIS web interface. On the left, there is a search bar with the query '*Czech*' and a corpus list showing 'GUM' with 126 texts and 109,140 tokens. The main area displays search results for 'Czech', showing a syntax tree diagram for the sentence 'and were intended to convey Czech national spirit, as were'. The tree includes nodes like 'cc', 'aux', 'pass', 'madv', 'acompl', 'advcl', 'prep', 'obj', 'amod', and 'nsubj'. Below the tree, there is a text snippet about Antonin Dvořák, mentioning his birth in 1841 and his work as a composer, with various linguistic annotations like 'advcl', 'prep', 'obj', 'amod', and 'nsubj'.

Abbildung 34: ANNIS

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal), Browseranwendung
Aktivierung	Installation
Betriebssystem	
Benötigte Anwendungen	Pepper (Konverter), Salt (API)
Lizenztyp	Open Source (CC BY NC SA 4.0)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	elementar

Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA, discA, toMap, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, Markup, WIMP
Dateiformate	ANNIS format
Import	xml
Export	ANNIS format
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	schwer
Link	https://corpus-tools.org/annis/

4.3.2. AntConc

AntConc ist ein plattformübergreifendes Open-Source-Werkzeug zur Durchführung einfacher korpuslinguistischer Analysen von Texten. Es beinhaltet verschiedene Tools zur Suche und Darstellung von Konkordanzen und Kollokationen, Clustern und Wortlisten.

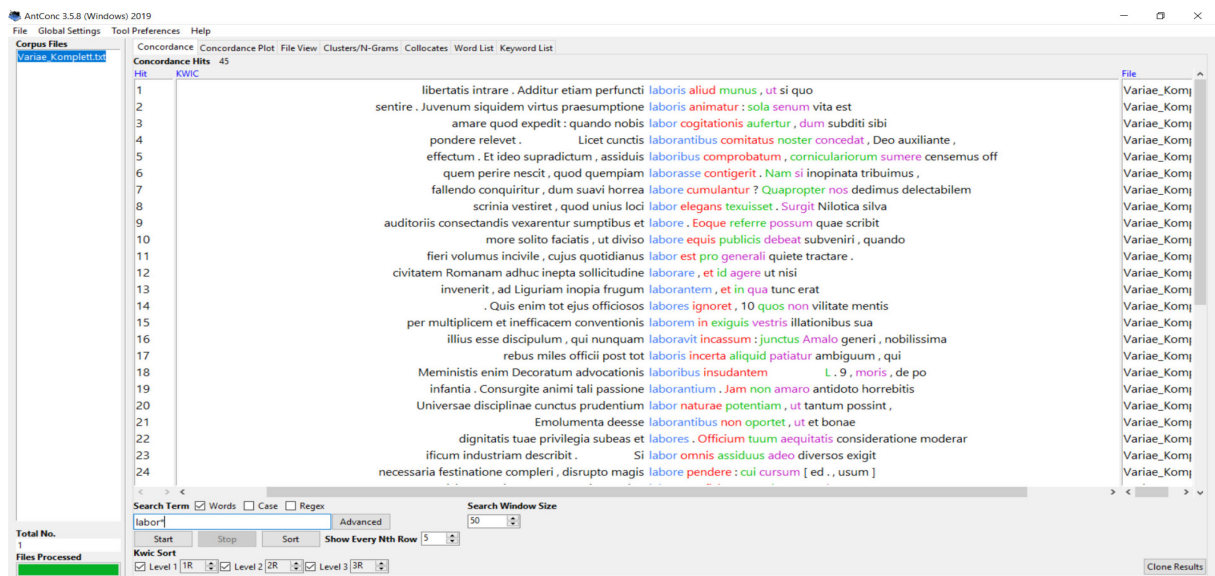


Abbildung 35: AntConc

Technische Daten	
Art der Anwendung	Portable Software
Aktivierung	Ausführbare JAR-Datei
Betriebssystem	Win, Linux, MacOS
Benötigte Anwendungen	
Lizenztyp	Freeware
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, Vis
Benötigte IT/DH-Vorkenntnisse	

Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	ant
Import	txt, xml, ant
Export	txt, x+AN7ml, ant, ps
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://www.laurenceanthony.net/software/antcon/

4.3.3. CorpusExplorer

Der CorpusExplorer ist eine Anwendung für korpuslinguistische und semantische Analysen von Texten und ermöglicht die vollautomatisierte Durchführung von Routineschritten zur Akquise und Annotation von Texten. Das Programm vereint in einer interaktiven Nutzeroberfläche zahlreiche Funktionen u.a. zur Analyse von Wortfrequenzen und Wortverteilungen, N-Gramme, KWIC, Phrasen, Stilmetriken und diskursanalytischen Gesichtspunkten sowie zur graphischen Aufbereitung der Resultate. Über die Software lassen sich Daten verschiedener Standardformate wie XML, CSV, JSON und R bearbeiten, auswerten und zur Weiterbearbeitung in anderen Anwendungen exportieren. Darüber hinaus bietet der Corpus Explorer ein Modul zur Entwicklung von Erweiterungen der Software oder auch zur Integration einzelner Funktionen in andere Programme.

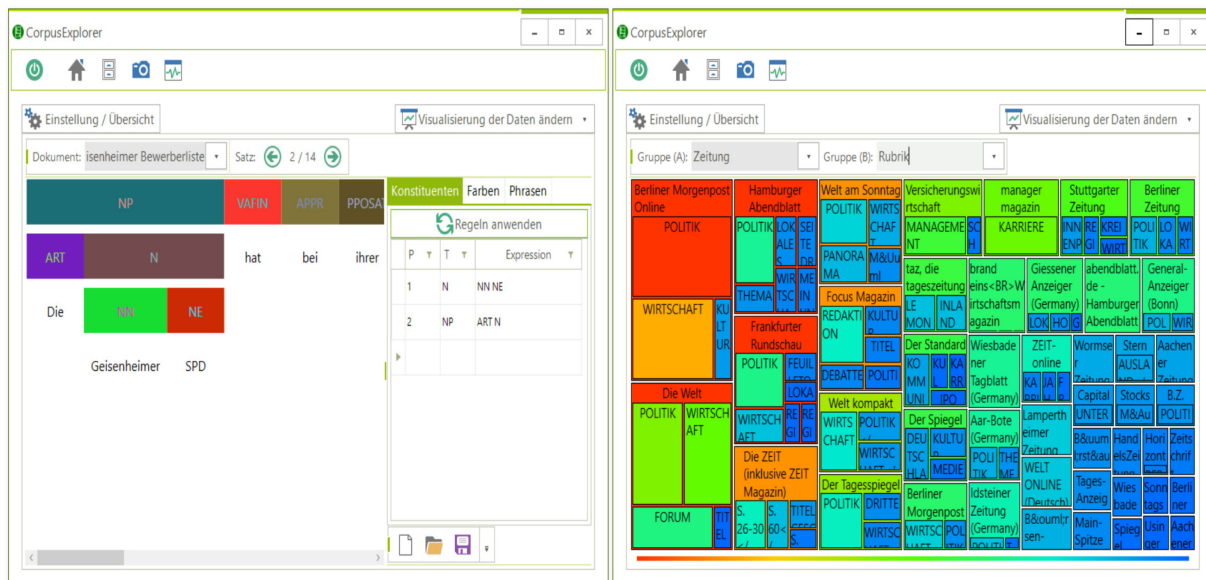


Abbildung 36: CorpusExplorer

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, Linux, MacOS
Benötigte Anwendungen	
Lizenztyp	Freeware

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal), Desktopanwendung (remote)
Aktivierung	Installation
Betriebssystem	MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v2), Freeware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA, discA
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt, cqp
Import	txt, cqp
Export	txt, cqp
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://cwb.sourceforge.net/

4.3.5. Cosmas II (Corpus Search, Management and Analysis System)

Cosmas ist ein umfangreiches Korpusrecherche- und Korpusanalysesystem, das sowohl als lokale Applikation als auch als Webanwendung genutzt werden kann. Es ermöglicht SprachwissenschaftlerInnen den Zugriff auf über 500 Korpora mit fast 60 Mrd. laufenden Wortformen. Die Recherche beinhaltet die Suche nach Wörtern und Wortteilen, Grundformen, Wortklassen, Textbereichen und Positionen sowie Wort- und Satzabständen und erlaubt die Differenzierung nach Entstehungszeit, Erscheinungsland, Thematik und Gebrauchskontexten. Dadurch können Kookkurrenzen, zeitliche Schwankungen im Gebrauch, seltene Gebrauchsformen und Neologismen identifiziert werden.

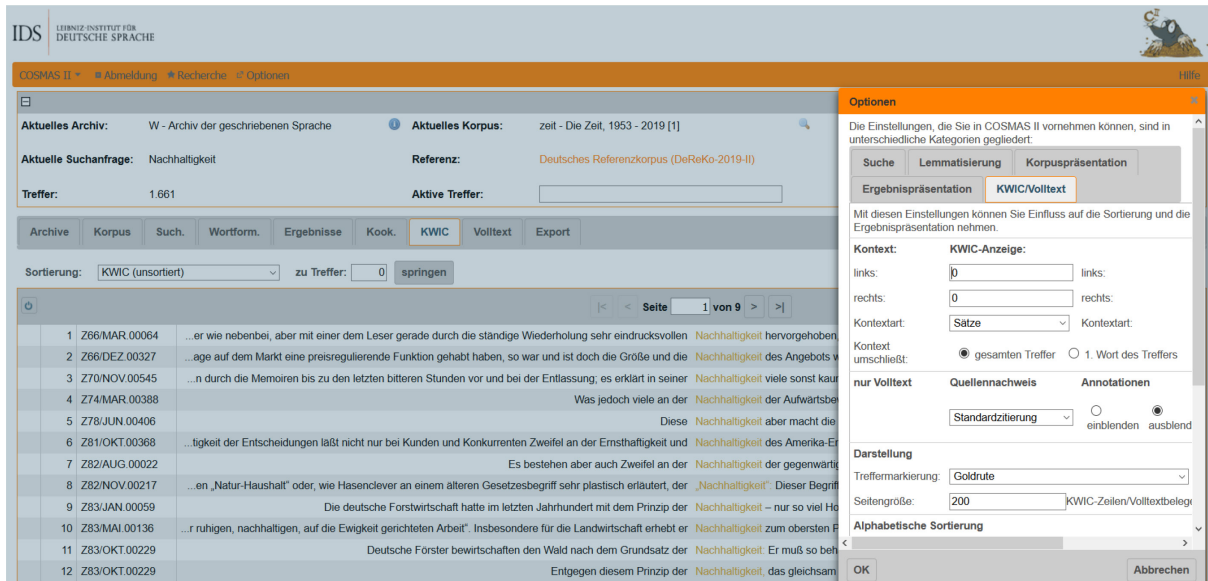


Abbildung 38: Cosmas II (Corpus Search, Management and Analysis System)

Technische Daten	
Art der Anwendung	Website, Desktopanwendung (lokal)
Aktivierung	Registrierung, Installation
Betriebssystem	Win
Benötigte Anwendungen	
Lizenztyp	Open Source (CC)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, sema, discA, tmpA
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP, Suchsyntax
Dateiformate	
Import	
Export	rtf, txt
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://cosmas2.ids-mannheim.de/

4.3.6. Hermetic Word Frequency Finder

Hermetic Word Frequency Counter führt Worthäufigkeitszählungen und Wortsuchen auf der Basis von docx-Scans durch. Über einen Konfigurator können einfache Wortklassifikationen festgelegt werden, die gefundenen Wörter alphabetisch oder nach Häufigkeit aufgelistet werden, wobei für jedes Wort Rang und Häufigkeitszahl angezeigt werden.

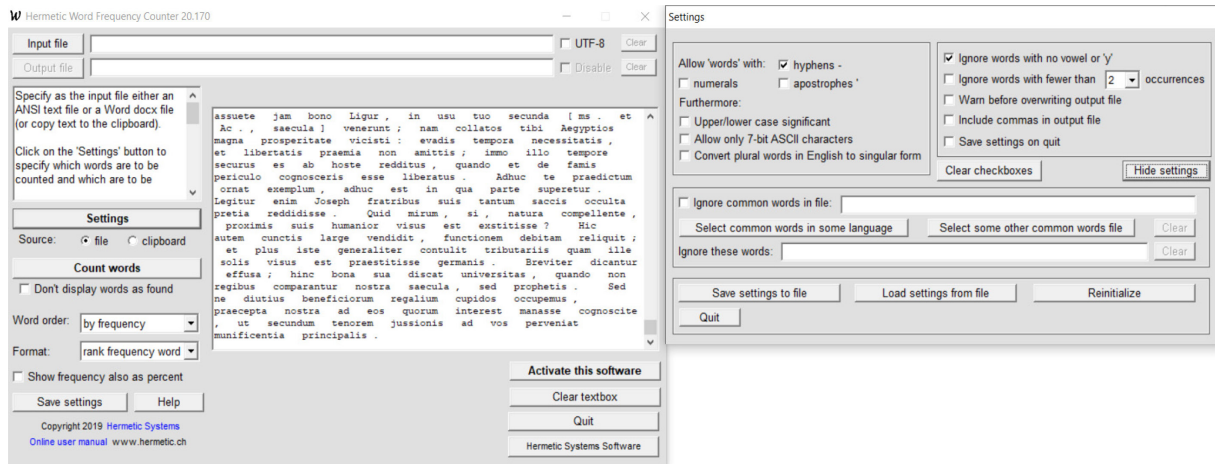


Abbildung 39: Hermetic Word Frequency Finder

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win
Benötigte Anwendungen	
Lizenztyp	Closed Source, Payware
Nutzerinnenunterstützung	
Dokumentation	begrenzt
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt
Import	txt, rtf, doc, weitere
Export	txt
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://www.hermetic.ch/wfc/wfc.htm

4.3.7. Jupyter Notebook

Jupyter Notebook ist eine browserbasierte interaktive Arbeitsumgebung, in der u.a. mittels Python-Skripten strukturierte Daten organisiert, bearbeitet, analysiert und visualisiert werden können. Das Programm erlaubt auf Basis minimalisierten Programmieraufwandes die Einbindung und Bearbeitung einer Vielzahl von Datenformaten wie Textdokumenten, Tabellen, Mediendaten und umfangreichen Datenbanken sowie deren Transformation und Auswertung. Die jeweils erstellten „Notizbücher“, in denen Daten und Verarbeitungsanweisungen miteinander verknüpft sind, können auf verschiedene Arten gespeichert und in Repositorien hinterlegt werden. Neben der Flexibilität in Bezug auf zu verarbeitende Datenformate wie txt,

rtf, xml, html oder tei unterstützt Jupyter Notebook eine Vielzahl von Programmiersprachen wie C#, PHP, JavaScript, Ruby und R.

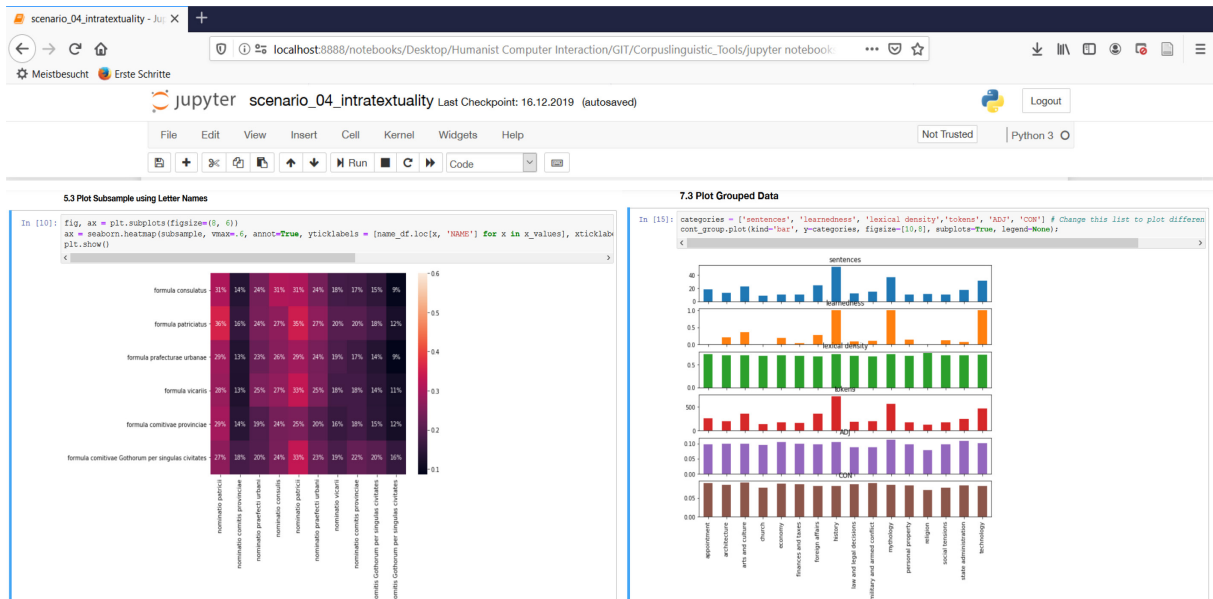


Abbildung 40: Jupyter Notebook

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	Python
Lizenztyp	Open Source (BSD 3Clause)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse, Netzwerkanalyse
Einzelmethoden	lingA, semA, discA, toMap, netwA, geoA, Vis
Benötigte IT/DH-Vorkenntnisse	Python
Eingabe und Formate	
Eingabemodi	WIMP
Dateiformate	ipynb, py
Import	py, ipynb, json
Export	ipynb, xml, html, latex, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	schwer
Link	https://jupyter.org/

4.3.8. JuxtaCommons

JuxtaCommons ermöglicht das Zusammenstellen und Vergleichen von Textversionen und bietet diverse Features zur Visualisierung der inhaltlichen und strukturellen Versionsunterschiede,

z.B. durch Side-by-Side-Maps oder Heatmaps. Kollationsergebnisse können auch als XML-Datei mit TEI-Segmentierung, als Word-Dokument oder HTML-Datei ausgegeben werden.

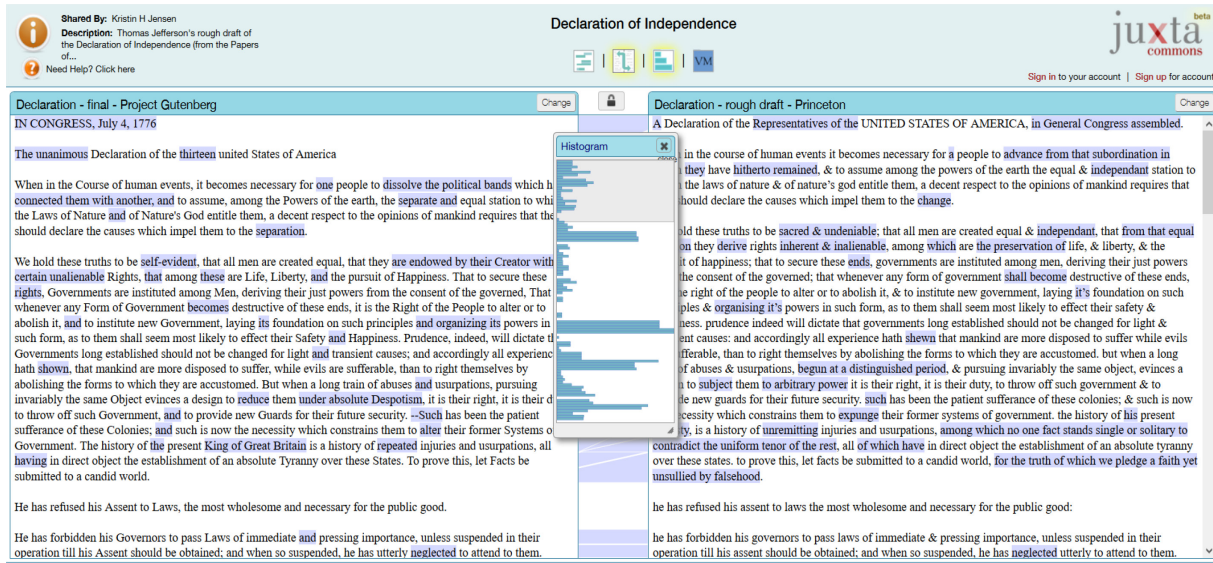


Abbildung 41: JuxtaCommons

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (Apache v2.)
NutzerInnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt, xml
Import	tei, txt, xml, html, weitere
Export	tei, xml, html
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	http://juxtacommons.org/

4.3.9. LancsBox (GraphColl)

LancsBox ist ein Werkzeug zur Erstellung und Analyse von linguistischen Kollokationsnetzwerken. Zur Untersuchung semantischer Zusammenhänge in Texten und Textkorpora werden diese innerhalb bestimmter Diskursfelder in Form lexikalischer Muster organisiert und die daraus resultierenden Kollokationsstrukturen in Form von Wortnetzwerken visualisiert. Das

Programm wird lokal installiert und bietet auf einer graphischen Benutzeroberfläche neben Standardeinstellungen die volle Kontrolle über statistische und methodische Funktionen zum Aufbau von Kollokationsnetzwerken. Das Interface ist um verschiedene assistierende Registerkarten aufgebaut, über die Kollokationsgraphen erstellt, erforscht und exportiert werden können. NutzerInnen können Eigenschaften des Kollokationsgraphen wie z.B. die Spanne der Kollokationsfenster, die Assoziationsmaße, die Kollokationsfrequenz und Boolesche Schwellenwerte festlegen.

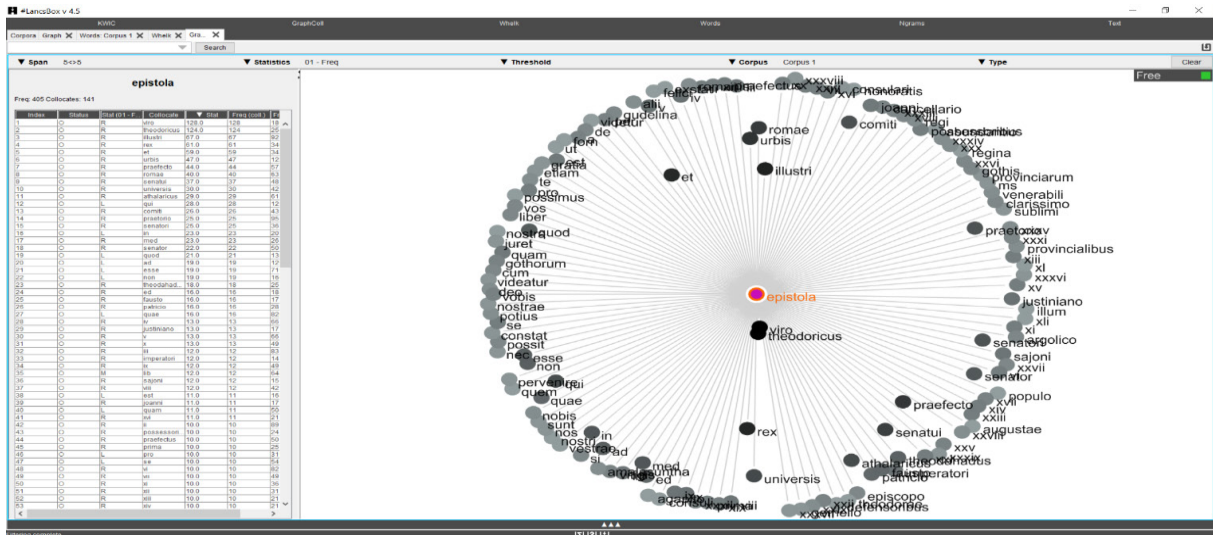


Abbildung 42: LanCSBox (GraphColl)

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Ausführbare JAR-Datei
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (CC BY NC ND)
NutzerInnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse, Netzwerkanalyse
Einzelmethoden	lingA, semA, discA, netwA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txt, xml, tei, html
Import	txt, xml, tei, html
Export	txt, png
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	http://corpora.lancs.ac.uk/lancsbox/

4.3.10. Rapid Miner

Rapid Miner ist ein Data-Mining-Programm, das die Erstellung, Bearbeitung, Analyse und Visualisierung vorstrukturierter Textdaten erlaubt. Das Programm bietet umfassende Anpassungsmöglichkeiten beim Datenimport, um Analysen gezielt vorzubereiten. Über ein umfangreiches Visualisierungstool lassen sich statistische Abfragen von Häufigkeiten, Verteilungen und Beziehungen zwischen Daten darstellen. Darüber hinaus können die Dateien mit Annotationen versehen werden.

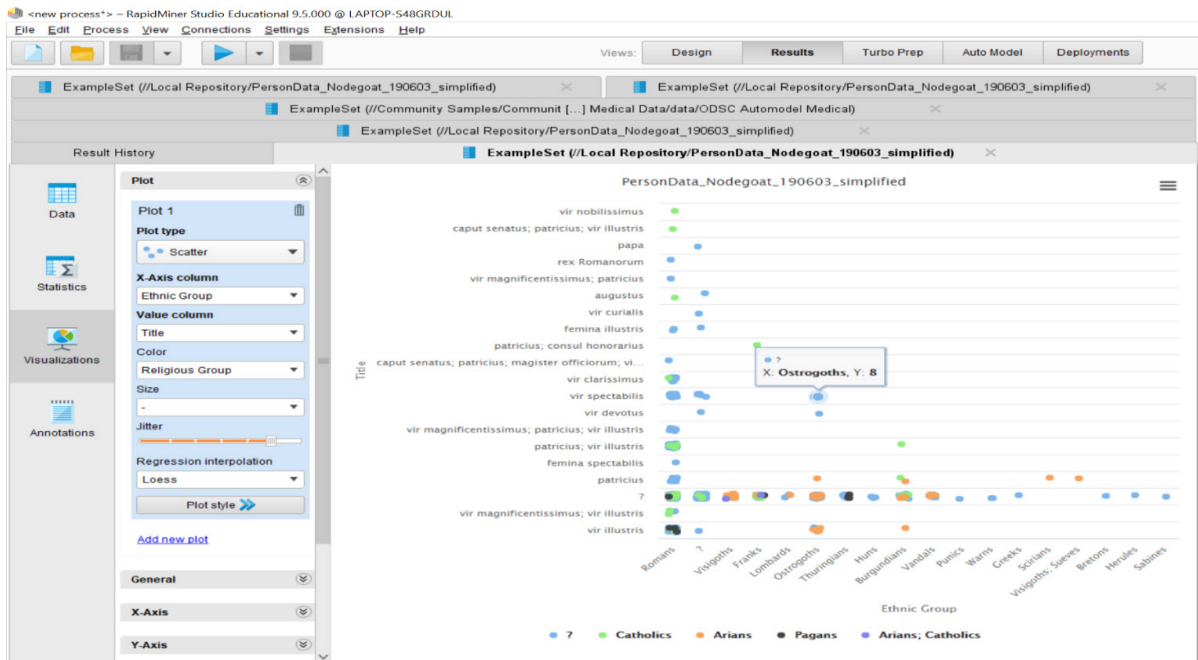


Abbildung 43: Rapid Miner

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (AGPL3.), Freeware, Payware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	semA, discA, toMap, netWA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	ioo, prop
Import	csv, xls, txt
Export	csv, xls
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://rapidminer.com/products/studio/

4.3.11. RAW Graphs

RAW Graphs ist eine Webanwendung auf Grundlage eines Open-Source-Datenvisualisierungs-Frameworks, das mit dem Ziel entwickelt wurde, die visuelle Darstellung komplexer Daten einfach zu gestalten. Das Tool bietet zahlreiche Templates für multidimensionale Diagramme, Plots und Charts, die in Abhängigkeit von der Struktur der Quelldaten zur Visualisierung genutzt und über ein Einstellungsmodul weiter angepasst werden können.

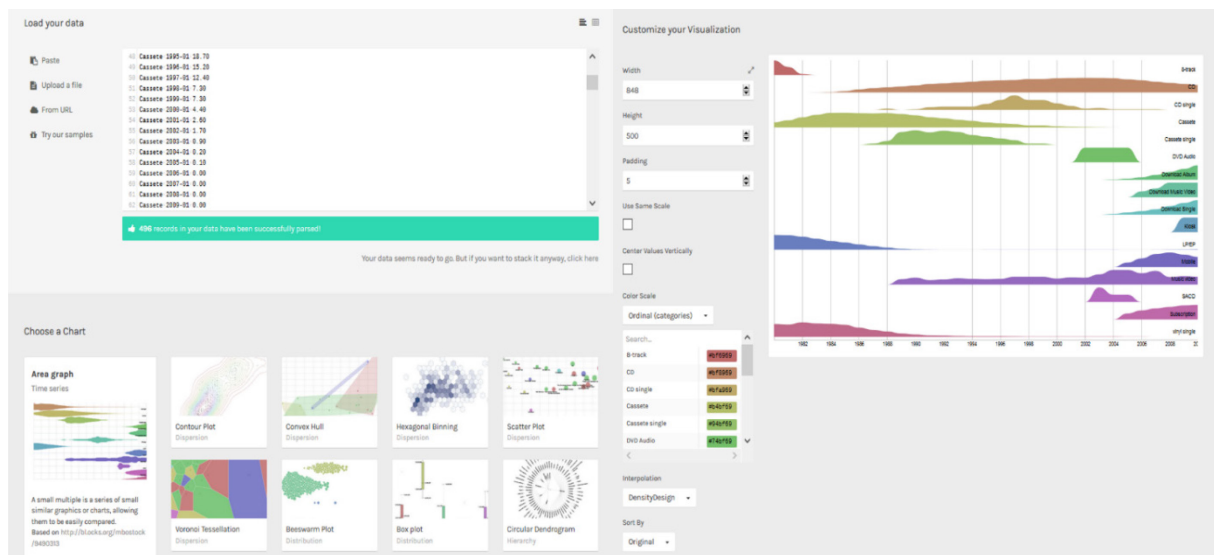


Abbildung 44: RAW Graphs

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (Apache 2)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	semA, toMap, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	WIMP
Dateiformate	csv
Import	csv, json, dsv, tsv
Export	json, png, jpg
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	https://rawgraphs.io/

4.3.12. Sketch Engine (NoSketch Engine)

Sketch Engine ist ein Werkzeug zur Sprachanalyse, für Text-Mining und Textanalyse. Es unterstützt die Untersuchung typischer oder ungewöhnlicher linguistischer Phänomene genauso wie lexikographische, semantikbezogene oder diskursanalytische Fragestellungen. Es erlaubt die Durchführung von Kookkurrenz- und Konkordanzanalysen, N-Gramm- und Worthäufigkeits-Analysen, die Recherche von Synonymen sowie wort-, phrasen- und textbezogene Kontextabfragen. Neben einem Fundus von 500 Korpora in 90 Sprachen mit 30 Mrd. Wortformen können eigene Korpora untersucht werden. Die Anwendung ist nur in der Testversion und als Desktopvariante NoSketch Engine kostenfrei.

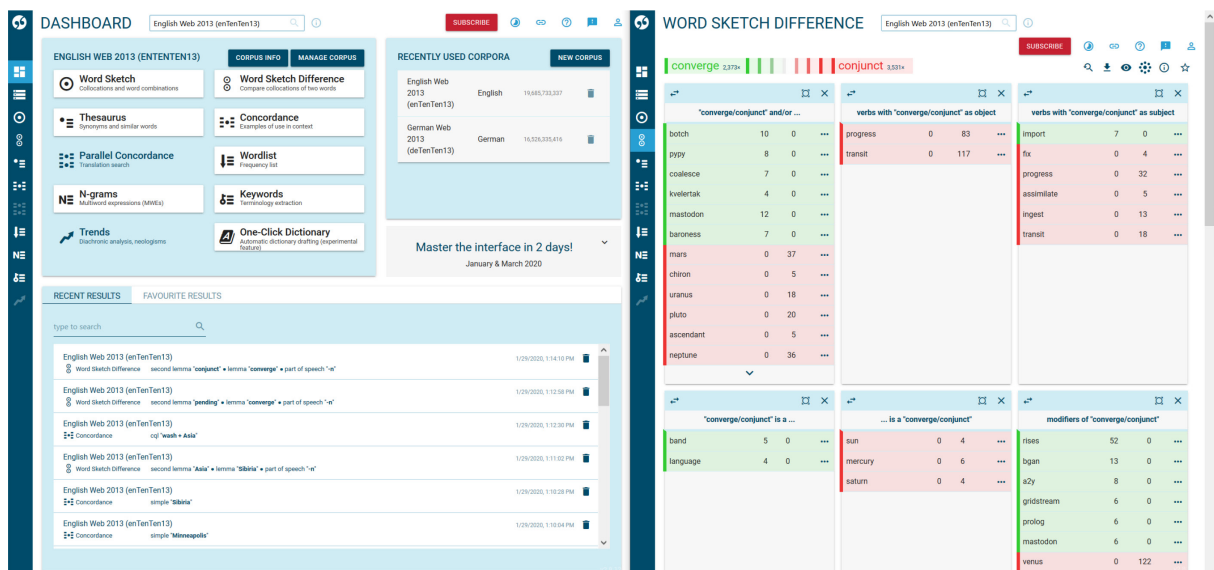


Abbildung 45: Sketch Engine (NoSketch Engine)

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL), Payware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, sema, disca, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	txt, html, xml, csv, xls
Export	txt, xml, doc, html, tei, zip
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat

Struktur der Steuerelemente Bedienung	übersichtlich leicht
Link	https://www.sketchengine.eu/

4.3.13. Stereoscope

Stereoscope ist eine webbasierte Anwendung zur Visualisierung von Textbedeutungen und Argumenten in Texten. Es erlaubt die visuelle Darstellung digitaler Annotationen, um qualitative Ergebnisse über den Gehalt und den argumentativen Aufbau von Texten gewinnen zu können. Unterschiedliche visuelle Darstellungen von Annotationsdaten erlauben verschiedene Perspektiven, die in Kombination mit Kommentaren und Tags Möglichkeiten zur wissenschaftlichen Beschäftigung mit literarischen Texten bereitstellen und diese für LeserInnen transparent machen.

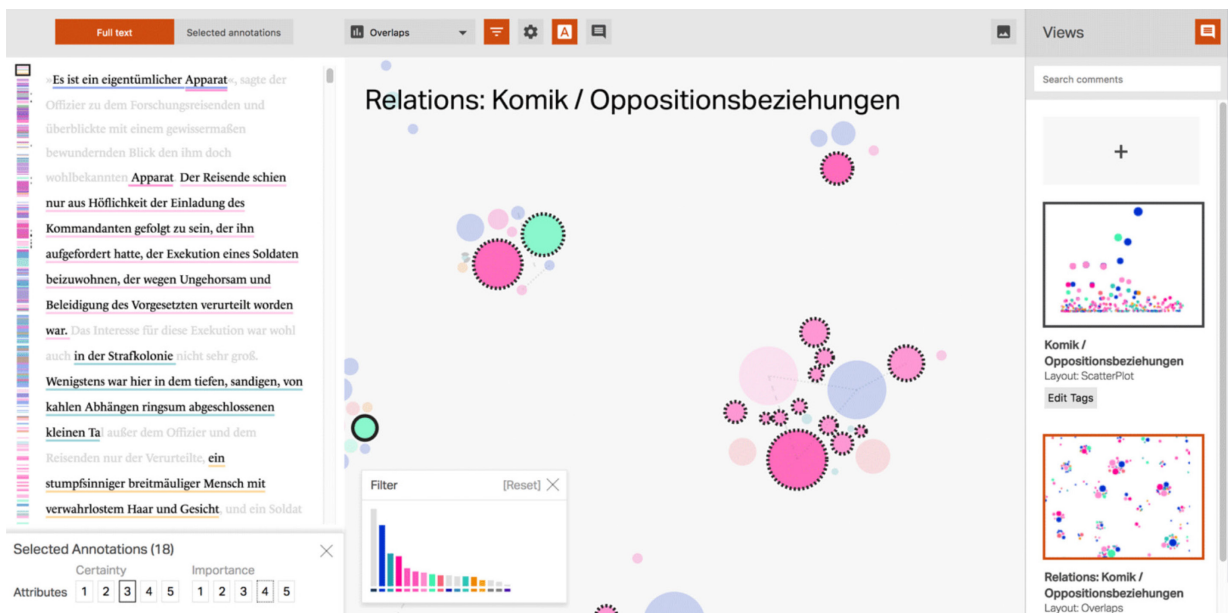


Abbildung 46: Stereoscope

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	Catma
Lizenztyp	Open Source
Nutzerinnenunterstützung	
Dokumentation	begrenzt
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse, Netzwerkanalyse
Einzelmethoden	lingA, sema, discA, toMap, sentA, netwA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	json
Import	json, txt

Einzelmethoden Benötigte IT/DH-Vorkenntnisse	lingA, semA, discA, netwA, Vis Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP txt txt, tei, html txt, pdf, png, jpg, svg
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	funktionsorientiert hoch differenziert anspruchsvoll
Link	https://fortext.net/tools/stylo

4.3.15. TAPORware

TAPORware ist eine Sammlung von Werkzeugprototypen, die zur text- und sprachanalytischen Analyse von HTML, XML und Reintextdateien genutzt werden können. Die Tools lassen sich zur Analyse sowohl von Webseiten als auch lokaler Dateien einsetzen, werden ohne Support frei zur Verfügung gestellt und sind nicht systematisch getestet worden. Zu jedem Werkzeug sind jedoch kleinere Hilfsanweisungen verfügbar.

The screenshot shows the TAPORware web interface. On the left is a navigation menu with categories like HTML Tools, List words, and Plain Text Tools. The main area is titled 'Tokenize Text with Specified Separators'. It has a 'Source text' section with a URL input field and a 'Token types' section with radio buttons for Words, Lines, Sentences, Paragraphs, and Regular expressions. Below that are 'Results' options for display and separator. On the right, there is a table of token counts and a bar chart showing the distribution of token counts.

Abbildung 48: TAPORware

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Closed Source, Freeware
Nutzerinnenunterstützung	
Dokumentation	keine
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA, Vis
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP

Dateiformate Import Export	txt, xml, html
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	funktionsorientiert hoch übersichtlich moderat
Link	http://taporware.ualberta.ca/

4.3.16. Textable

Textable ist ein Add-on-Tool, das für das Orange-Data-Mining- und -Visualisierungspaket entwickelt wurde. Es erlaubt den Aufbau von Datenmodellen und Datentabellen aus Textdaten, den Import von Daten aus anderen Quellen und die Analyse von Konkordanzen und Kollokationen. Darüber hinaus können BenutzerInnen Texte segmentieren und kommentieren, Daten systematisch umcodieren, Teilmengen bestimmen und Frequenz- und Komplexitätsmessungen durchführen.

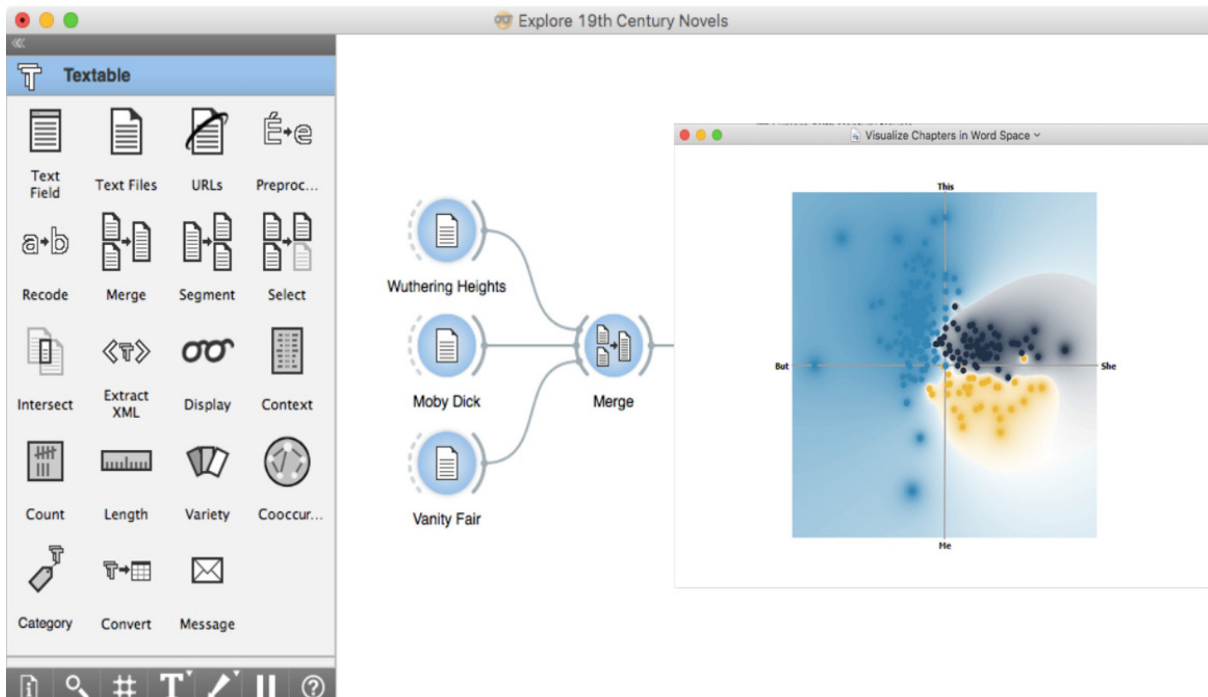


Abbildung 49: Textable

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	Orange Canvas
Lizenztyp	Open Source (GPL 3., LGPL 2.1)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar

Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, semA, discA, toMap, sentA, netWA, Gmod, Vis
Benötigte IT/DH-Vorkenntnisse	Datenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	tab
Import	csv
Export	tab, csv, png
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	anspruchsvoll
Link	http://textable.io/

4.3.17. TextDNA

TextDNA erlaubt die Wortanalyse in Rohdaten und Textsammlungen unterschiedlicher Größenordnung sowie den Vergleich von Dokumentsammlungen, Dokumenten oder einzelnen Elementen innerhalb eines Dokuments. Über verschiedene Metriken können absolute und relative Worthäufigkeiten, Wortfolgen, Wortklassen und Wortverwendungsmuster identifiziert werden. Das Programm arbeitet auf Basis des Genom-Analysesystems Sequence Surveyor, wobei Wörter als Datensequenzen behandelt und dementsprechend durch farbliche und positionale Aggregation visualisiert werden. Die Sequenzkolorierung ermöglicht die Identifikation von globalen Mustern und Sonderfällen in Texten.

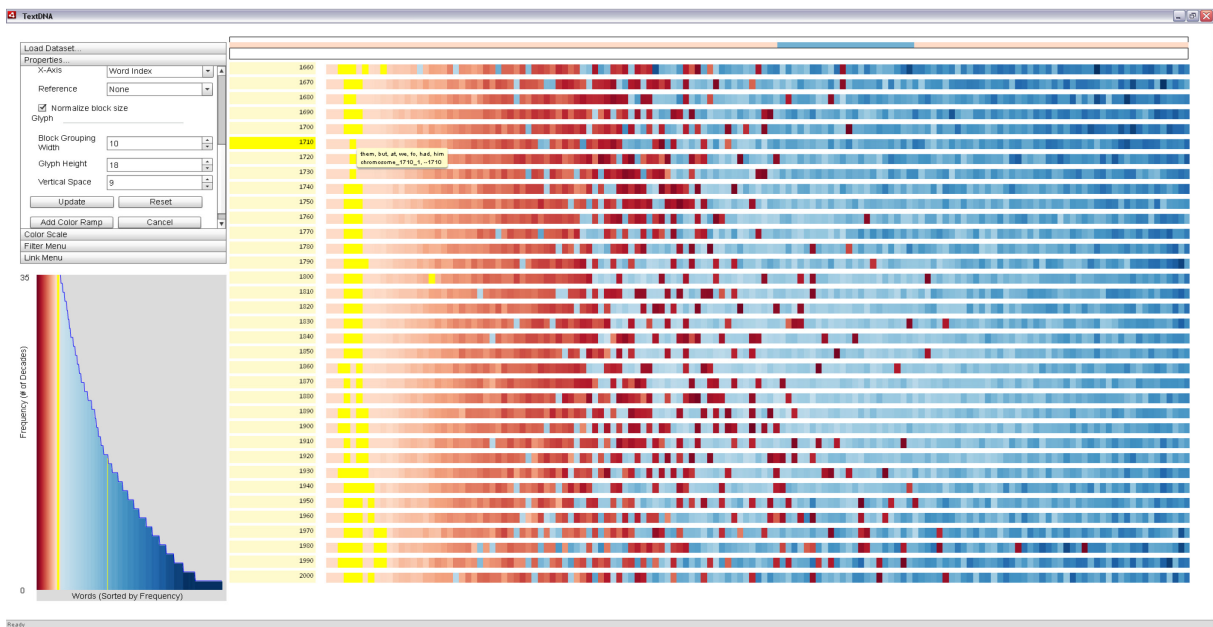


Abbildung 50: TextDNA

Technische Daten	
Art der Anwendung	Browseranwendung
Aktivierung	
Betriebssystem	

Benötigte Anwendungen Lizenztyp	XAMPP
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	elementar nein nein nein
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Korpuslinguistische Analyse semA, discA, toMap, sentA, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert moderat einfach moderat
Link	http://graphics.cs.wisc.edu/Vis/SequenceSurveyor/TextDNA.html

4.3.18. TokenX

TokenX ist ein Webtool für Wortzählungen, Wortwolken und einfachen Wortstatistiken.

Abbildung 51: TokenX

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizenztyp	Website Freeware
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	keine nein nein nein

Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Korpuslinguistische Analyse lingA, semA, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP xml xml
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert gering einfach leicht
Link	http://tokenx.unl.edu/

4.3.19. Tropes

Tropes ist ein Textanalysetool und Thesaurusmanager zur Verarbeitung natürlicher Sprache in Form semantischer Klassifizierungen, einschließlich der chronologischen Analyse von sequentiellen Texten und ihrer thematischen Zusammenfassung. Es bietet praktische Hilfsmittel zur Überprüfung textanalytischer Hypothesen und deren Interpretation. Die Tropenanalyse erfolgt über eine graphische Benutzeroberfläche, die automatisch sortierte Keywordlisten erstellt und eine Vielzahl korpusanalytischer Visualisierungen erlaubt. Über die verschiedenen Klassifizierungen wie die Relation von Sprechenden und Adressierten, Handlungssituationen und Zwecken können relevante Textinformationen extrahiert und somit Zusammenhänge erkannt, Themen isoliert und AkteurInnen identifiziert werden. Die qualitative Analyse arbeitet mit semantischen Metakategorien, um Verben, Adjektive, Adverbien, Personalpronomen und Konjunktionen zu gruppieren, Textstile zu identifizieren und Kontextrelationen zu anderen Texten aufzuzeigen. Durch Isolation einzelner Textepisoden, der Visualisierung von Themensträngen und der Häufigkeit von Konzepten im Textverlauf, lassen sich chronologische Analysen durchführen. Des Weiteren können durch Identifikation fundamentaler Propositionen im Text auch relevante Informationen für die Diskursanalyse gewonnen werden.

integrierten Sprachverarbeitungswerkzeugen (TAL), wie dem Morphosyntax Labeler und dem TreeTagger Lemmatizer. Die graphische Benutzeroberfläche bietet vielfältige Funktionen zur Korpusaufbereitung wie die Isolierung und Partitionierung von Subkorpora aus verschiedenen eigenschaftsbezogenen Metadaten von Texten, die Erstellung von HTML-Editionen, den Aufbau eines Korpus aus verschiedenen Textquellen und umfangreiche Importmodule zur Einbettung von Klartext, CSV-Daten, TEI/XML-Dokumenten sowie weiterer Textdateiformate. Auswertungsergebnisse können in CSV, XML oder SVG exportiert werden. Auf der analytischen Ebene sind intertextuelle Vergleiche, Konkordanz- und Kollokationsanalysen, statistische Wortzählungen auf Grundlage von Gesamttexten und bestimmter Wortheigenschaften, die Erstellung von Kontingenztabellen sowie verschiedene Tools zur Korpusvisualisierung nutzbar. Die Suchfunktion ermöglicht über den internen Corpus Query Processor (CQP) die Abfrage von Wörtern, Zeichenketten, Wortfolgen sowie grammatikalischen Kategorien. Daher eignet sich TXM auch für phraseologische Analysen. Sich wiederholende oder aufwändige Arbeitssitzungen können durch Erstellung Groovy- und R-basierter Skripte reproduziert werden. Darüber hinaus lässt sich bei Bedarf mehr als ein Dutzend weiterer Zusatzanwendungen aus dem Entwicklungsumfeld des TEI-Konsortiums in der Anwendungsoberfläche integrieren.

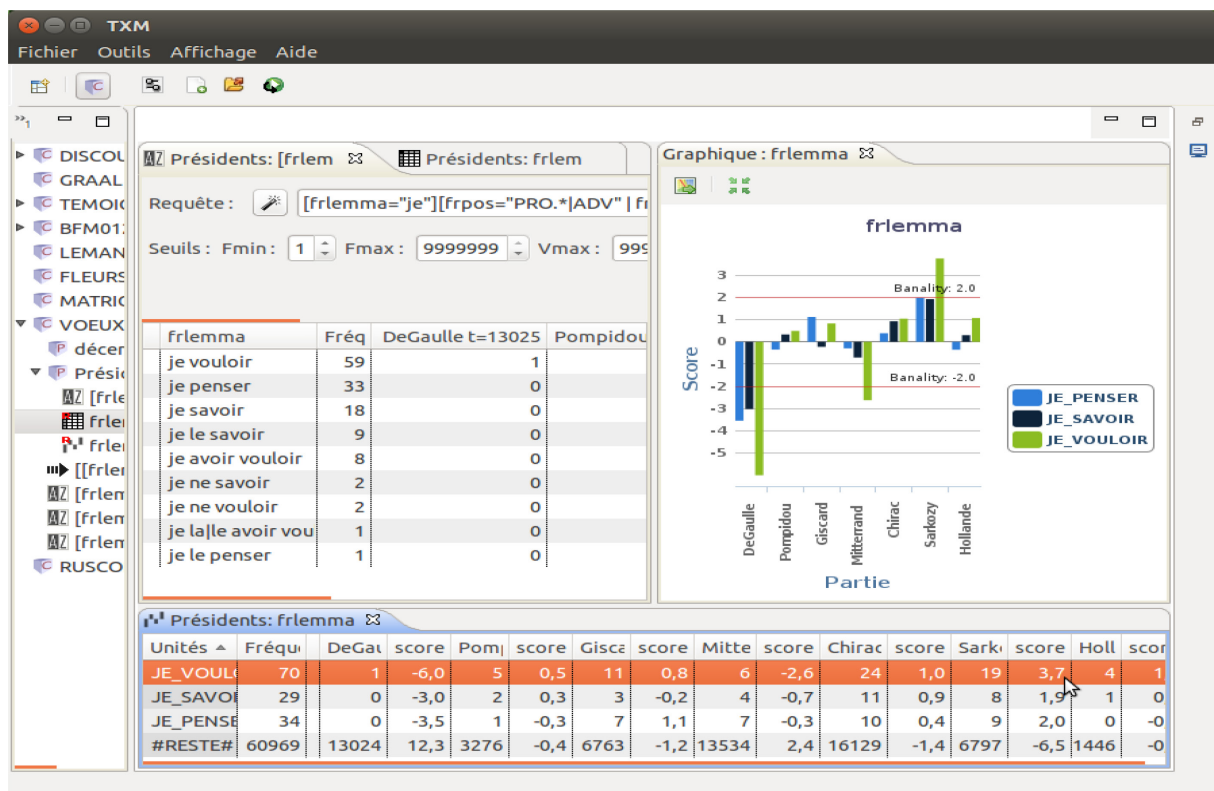


Abbildung 53: TXM (Textometrie)

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (GPL v2)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, sema, discA, toMap, Vis
Benötigte IT/DH-Vorkenntnisse	Computerlinguistische Grundlagen
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	txm
Import	xml, csv, cnr, txt
Export	txm, xml
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	schwer
Link	http://textometrie.ens-lyon.fr/?lang=en

4.3.21. Voyant Tools

Voyant ist eine virtuelle webbasierte Engine zur quantitativen Textanalyse und beinhaltet viele linguistische Analyse- und Visualisierungstools wie Häufigkeits- und Frequenzanalyse, Konkordanzanalyse und Wortwolken. Das Webtool ist auch in andere DH-Anwendungen wie z.B. TextGrid integriert.

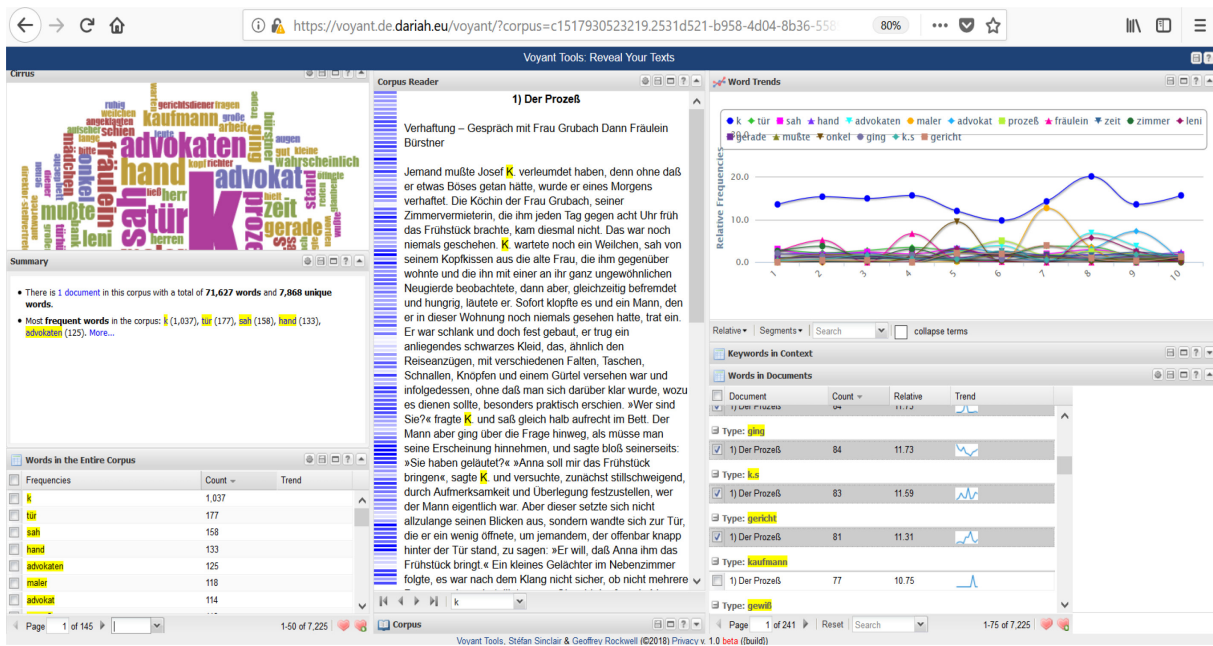


Abbildung 54: Voyant Tools

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizenztyp	Website Open Source (GPL v3.)
NutzerInnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	ausführlich elementar ja nein
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Korpuslinguistische Analyse lingA, sema, discA, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	WIMP txt, doc, xml, tei, weitere Textformate json, png
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert moderat übersichtlich leicht
Link	http://voyant-tools.org/

4.3.22. Word and Phrase

Word and Phrase ist ein frei zugängliches Webtool, das über die Schnittstelle des Corpus of Contemporary American English (COCA) Informationen zu den in Texten verwendeten Wortformen und Phrasen liefert. Es erlaubt die Analyse von Worthäufigkeiten, Kollokationen und Verwendungskontexten von Wortformen und Wortfolgen.

Abbildung 55: Word and Phrase

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizenztyp	Website Freeware
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	elementar nein ja nein
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Korpuslinguistische Analyse lingA, sema, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert moderat übersichtlich leicht
Link	https://www.wordandphrase.info/

4.3.23. WORDij

WORDij ist ein kostenfreies semantisches Netzwerktool mit verschiedenen Modulen zur automatisierten Verarbeitung natürlicher Sprache und der Analyse und Visualisierung unstrukturierter Texte aus beliebigen Textquellen. Das Programm ist in C geschrieben und nach einer Registrierung unter Windows, Mac und Linux installierbar. Fundament ist das WordLink-Modul, das Wörter und Wortpaare zählt, Beziehungen zwischen diesen erfasst und die verschiedenen Ergebnisse in separaten Dateien bzw. Dateiformaten ausgibt. Die enthaltenen Informationen sind Grundlage für weiterführende Analysen, wie z.B. die Abbildung der Wortnetzstruktur. Es können Textdateien in Bezug auf Unterschiede von Wörtern und Wortpaaren verglichen, Ähnlichkeiten zwischen Wortnetzwerken analysiert, Wortnetzwerke für einen oder mehrere Texte graphisch visualisiert, semantische Konstrukte hervorgehoben und Substantive extrahiert werden. WORDij-Ausgabedateien lassen sich in andere Netzwerkanalyseprogramme wie UCINET, NodeXL, Pajek oder Negopy importieren und dort weiterbearbeiten.

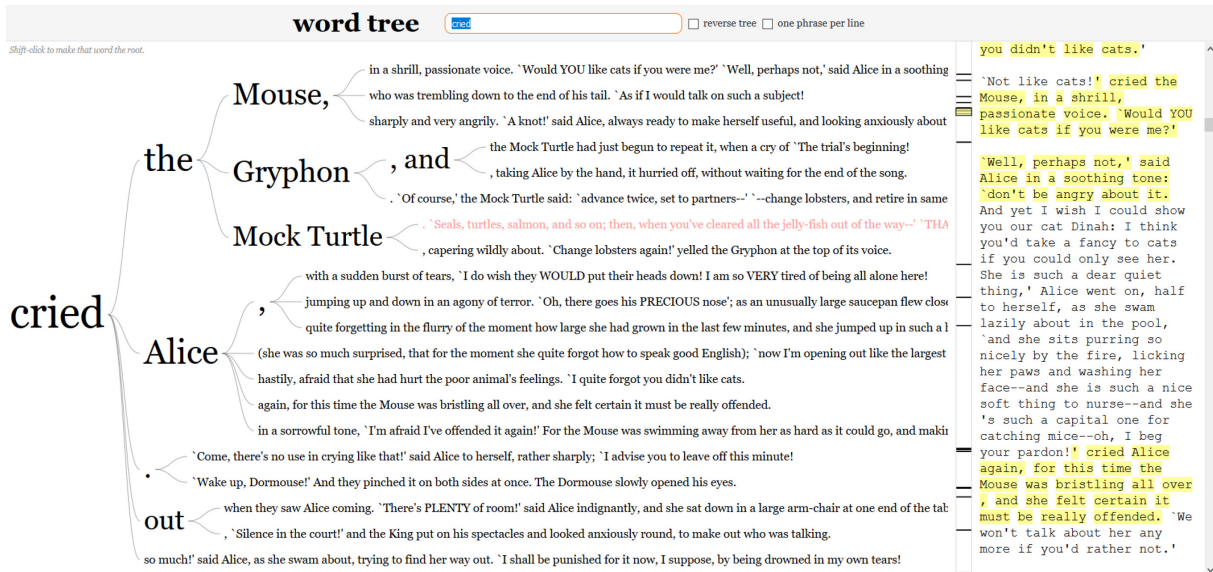


Abbildung 58: Wordtree

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Freeware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Korpuslinguistische Analyse
Einzelmethoden	lingA, discA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://www.jasondavies.com/wordtree/

4.4. Tools zur Netzwerkvisualisierung

4.4.1. Cytoscape

Cytoscape ist eine Open Source Software zur detaillierten Netzwerkanalyse und -visualisierung. Ursprünglich zur Visualisierung biomolekularer und genetischer Pfadannotationen entwickelt, hat es mittlerweile auch Anwendung in der linguistischen und sozialwissenschaftlichen Netzwerkanalyse gefunden. Das funktionsreiche Programm ist auf die Arbeit mit webbasierten

Systemen ausgerichtet, kann aber für jede Art von Attributen verwendet werden. Neben einer Reihe grundlegender Module zur Datenintegration, -analyse und -visualisierung stehen Funktionen für die Analyse von Netzwerkprofilen, Layout, Skripting und die Verbindung mit Datenbanken zur Verfügung.

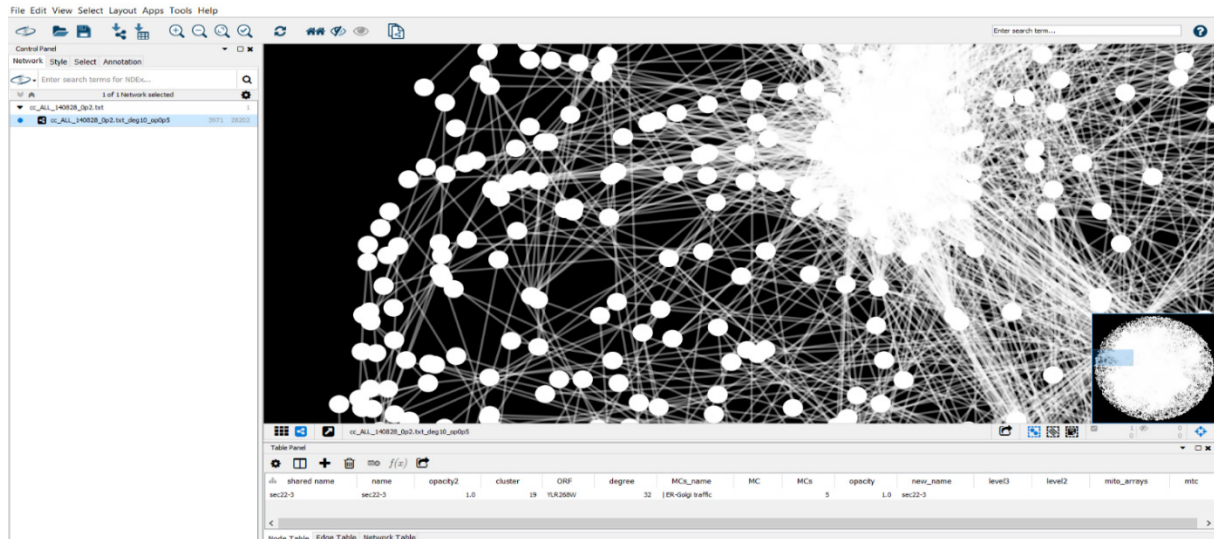


Abbildung 59: Cytoscape

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS
Benötigte Anwendungen	Java
Lizenztyp	Open Source (LGPL vv2.1)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	mAnn, semA, netwA, GMod, Vis, tmpA
Benötigte IT/DH-Vorkenntnisse	Graphdatenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	cys
Import	xml
Export	cx, cxjs, grapml, xml, sif, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	schwer
Link	http://www.cytoscape.org/

4.4.2. EgoWeb 2.0

EgoWeb 2.0 ist eine flexible Open-Source-Software zur Erfassung von Netzwerkdaten und erlaubt die Erstellung sowohl egozentrischer als auch kognitiver Netzwerke. Es ermöglicht verschiedene Arten der Datenerfassung und kann Daten visuell präsentieren. Exportierte Daten können in einer Reihe anderer Programme zur Netzwerkanalyse weiterverwendet werden.

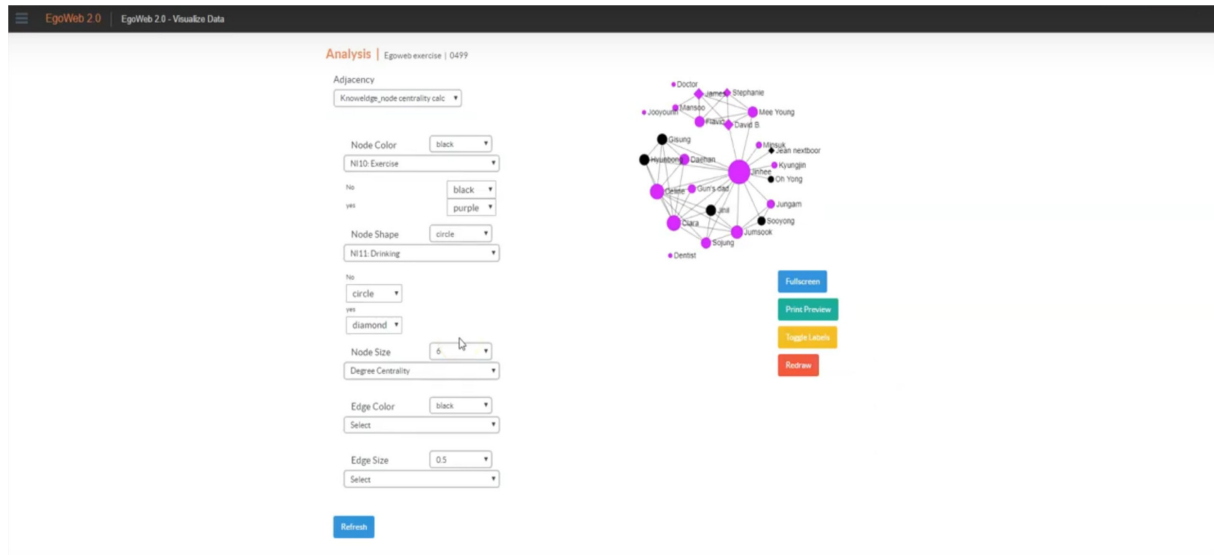


Abbildung 60: EgoWeb 2.0

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	Ampps, Java
Lizenztyp	Open Source (GPL)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	netwA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	study
Import	study, csv, xml, weitere
Export	study
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	https://www.qualintitative.com/egoweb/

4.4.3. Ezlinavis

Ezlinavis ist ein einfaches Werkzeug zur manuellen Eingabe relationaler Daten und der Echtzeitvisualisierung ungerichteter Netzwerke. Es eignet sich zur Repräsentation semantischer, linguistischer und personenbezogener Netzwerke anhand verschiedener Darstellungsalgorithmen. Darüber hinaus können die bearbeiteten Netzwerkdaten zur Weiterbearbeitung exportiert werden.

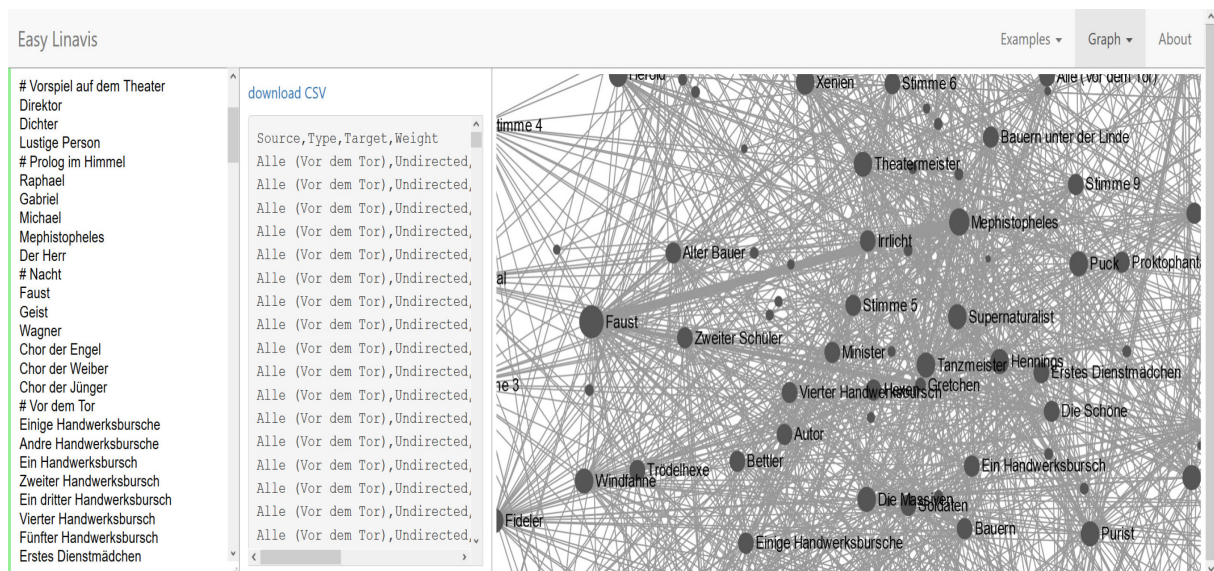


Abbildung 61: Ezlinavis

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Freeware
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	netwA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text
Dateiformate	
Import	
Export	csv, gexf
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://ezlinavis.dracor.org/

4.4.4. Gephi

Gephi ist eine Open-Source-Software zur Visualisierung und Analyse großer Netzwerkgraphen auf Basis einer 3D-Rendering-Engine, die Echtzeitvisualisierungen und detaillierte Analysen erlaubt. Es können praktisch alle Arten von strukturierten Daten als Netzwerke behandelt und visualisiert, gefiltert und geclustert, weiterbearbeitet und wieder exportiert werden. Neben vielen Analyseebenen wie Betweenness Centrality, Nähe, Durchmesser, Clusterkoeffizienten, PageRank, Modularität und Shortest Path bietet es zahlreiche Features für das Layout der Netzwerkvisualisierungen an. Darüber hinaus lassen sich auch zeitliche Veränderungen und geodatenbasierte Beziehungen analysieren und visualisieren. Anhand dynamischer Filter und Tabelleneditoren lassen sich Netzwerkdaten umfassend weiterbearbeiten.

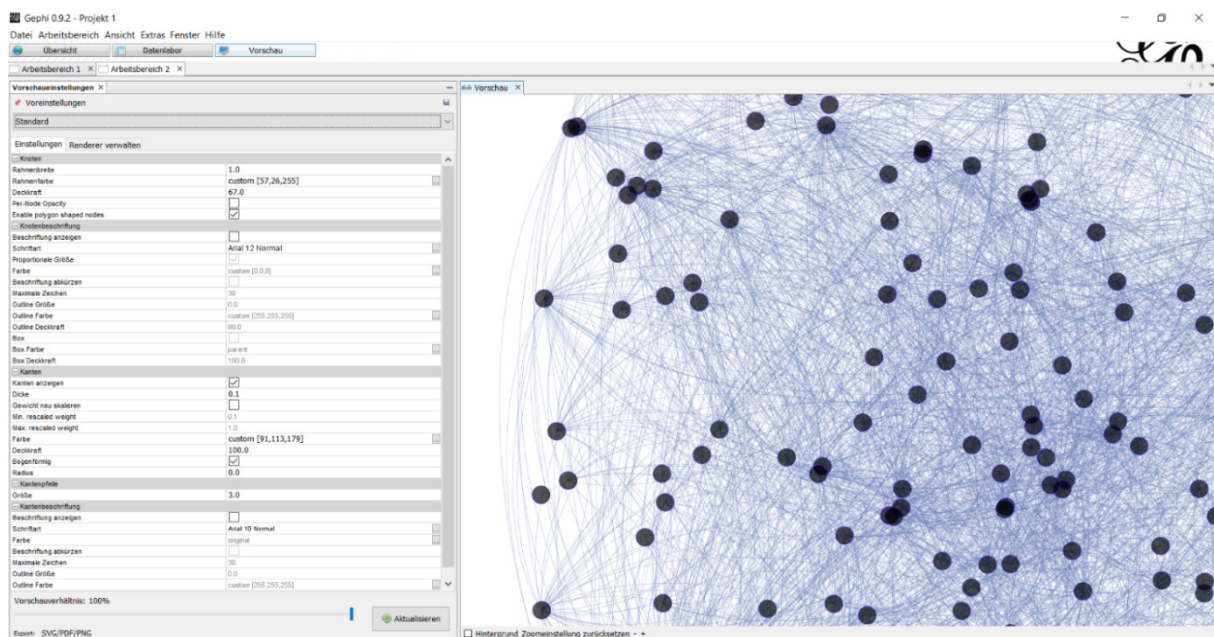


Abbildung 62: Gephi

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS
Benötigte Anwendungen	Java
Lizenztyp	Open Source (CDDL 1., GPL v3)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	netwA, GMod, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	Graphdatenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	gexf
Import	csv, graphml, gexf, gdf, net, gml, weitere
Export	gexf, graphml, gdf, gml, net, csv, weitere

Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	moderat
Link	https://gephi.org/

4.4.5. GraphViz

Graphviz ist eine Open-Source-Software zur Visualisierung von Graphen bzw. Netzwerken. Sie bietet eine Reihe von Funktionen für das Graphenlayout und die Formatierung von Diagrammen in einer einfachen Textsprache, über die Farben, Schriftarten, Knoten- und Kantenlayouts, Hyperlinks und benutzerdefinierte Formen festgelegt werden können. Neben der Einbindung und Modifikation externer Graphdaten können auch Netzwerke aus Rohtextdaten mittels eines graphischen Editors erstellt und weiterbearbeitet werden.

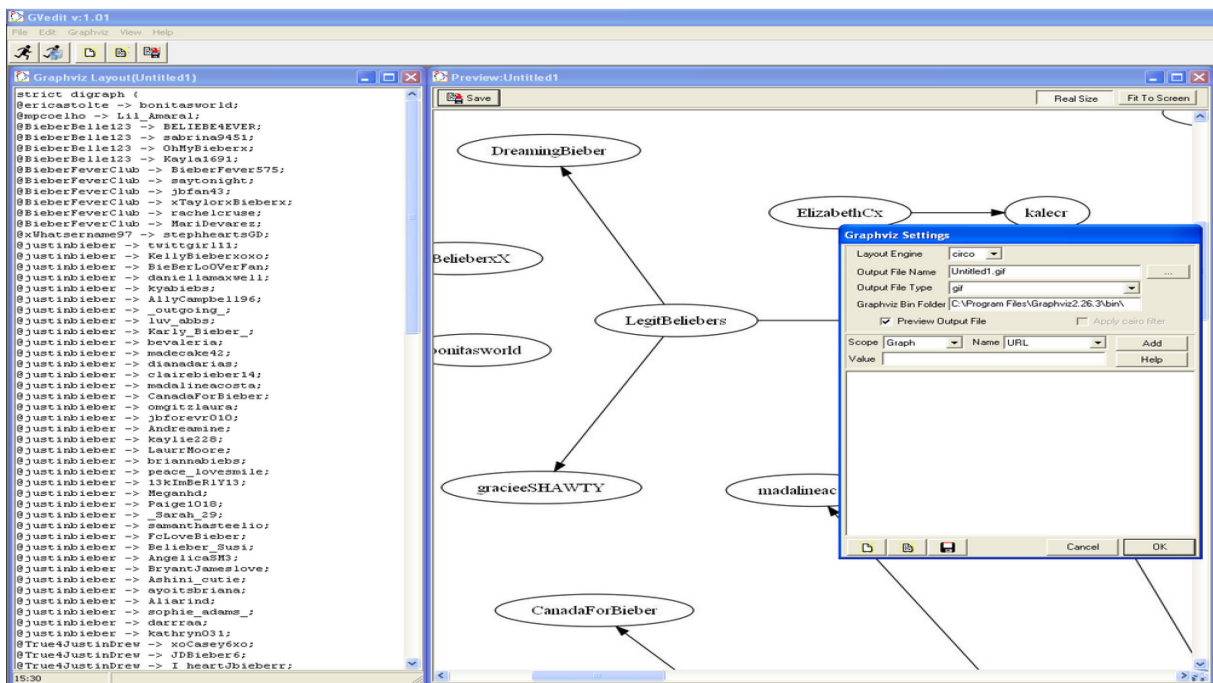


Abbildung 63: GraphViz

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source (EPL v1.)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	netwA, GMod, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	Graphdatenbanken

Eingabe und Formate	
Eingabemodi	Text, Markup, WIMP
Dateiformate	gxl, gml, xmd, gexf
Import	txt, gexf, gml, graphml, weitere
Export	png, svg, jpg, json, ps, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	differenziert
Bedienung	anspruchsvoll
Link	http://www.graphviz.org/

4.4.6. Histogramm

Histogramm ist eine frei zugängliche Webanwendung, die Netzwerke auf Basis von Kookkurrenzen in einer bereits vorhandenen multimedialen Datenbank von Personen, Bildern, Textdokumenten sowie Ort- und Zeitdaten erstellt. Unterstützend können Filter zur Eingrenzung des zu visualisierenden Dokumentenbereichs oder bestimmter Entitäten eingesetzt werden wie Datumsbereiche, Dokumentklassen oder auch bestimmte Wortgruppen, Tags und Normdaten. Die Ausgabe erfolgt in Form einer Liste und einer Netzwerkvisualisierung, in der die hinter den Knoten und Kanten liegenden Informationen interaktiv abgerufen werden können. Zur Bearbeitung eigener Daten ist eine neo4j-Instanz erforderlich.

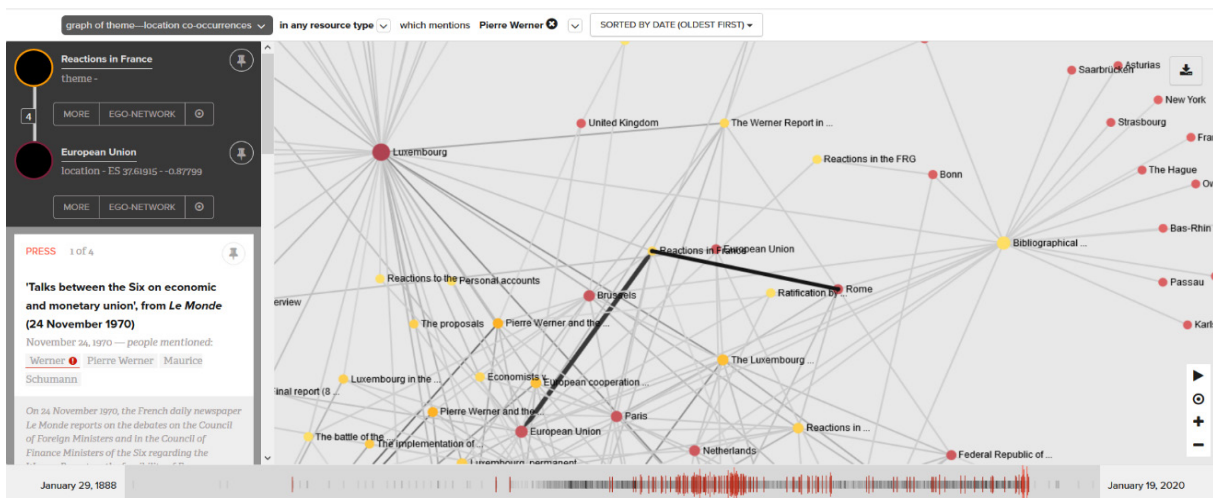


Abbildung 64: Histogramm

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse

Einzelmethoden	netwA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	WIMP
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	http://histograph.eu/

4.4.7. InfraNodus (Texttexture)

InfraNodus ist ein browserbasiertes Open-Source-Tool zur Untersuchung und Visualisierung textbezogener Netzwerke. Es eignet sich sowohl zur Identifikation von Schlüsselwörtern, Wortbeziehungen und Diskurselementen als auch zur Aufdeckung der Wortvielfalt oder struktureller Lücken in Textnetzwerken. Das Tool kann sowohl als eigenständige Software genutzt als auch über eine API in andere Tools implementiert werden.

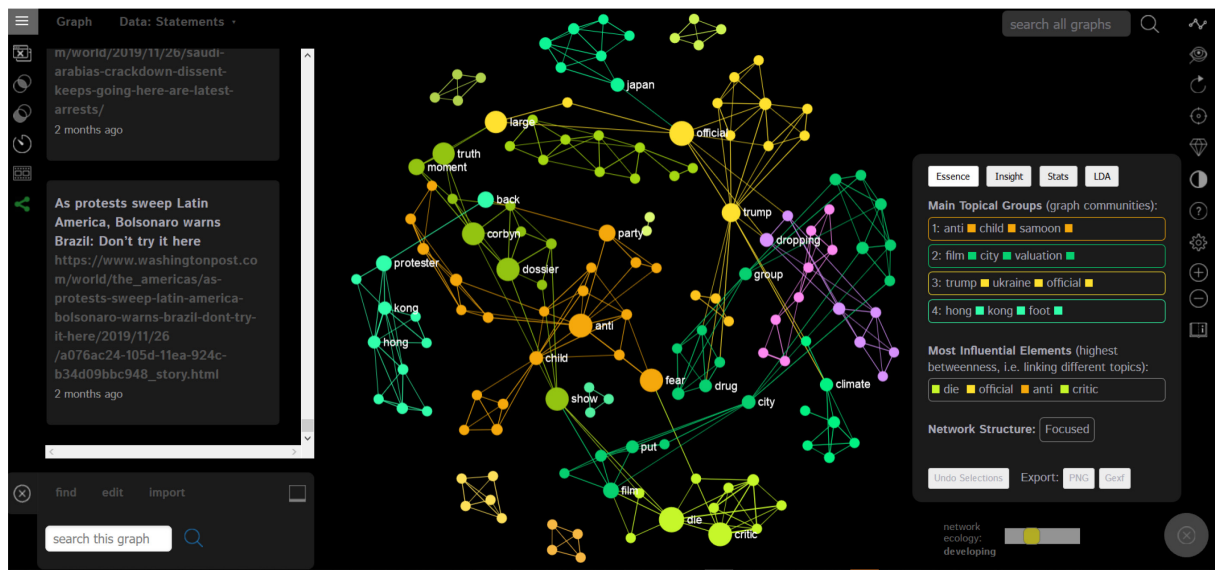


Abbildung 65: InfraNodus (Texttexture)

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal), Website
Aktivierung	Installation, Registrierung
Betriebssystem	Win
Benötigte Anwendungen	Neo4j, Java
Lizenztyp	Open Source (GPL), Payware (Webanwendung)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse

Einzelmethoden Benötigte IT/DH-Vorkenntnisse	netwA, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP neo4j db txt, gefx, weitere gefx, csv, png
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	funktionsorientiert moderat übersichtlich moderat
Link	https://infranodus.com/

4.4.8. Neo4j

Neo4j ist ein in Java implementierter Open-Source Datenbank-Editor, der Daten in Graphen strukturiert, speichert und visualisiert. Der Hersteller bietet ein Komplettpaket für die Datenanalyse inklusive Neo4j-Datenbank, Browser und einem Leitfaden mit Beispielen an.

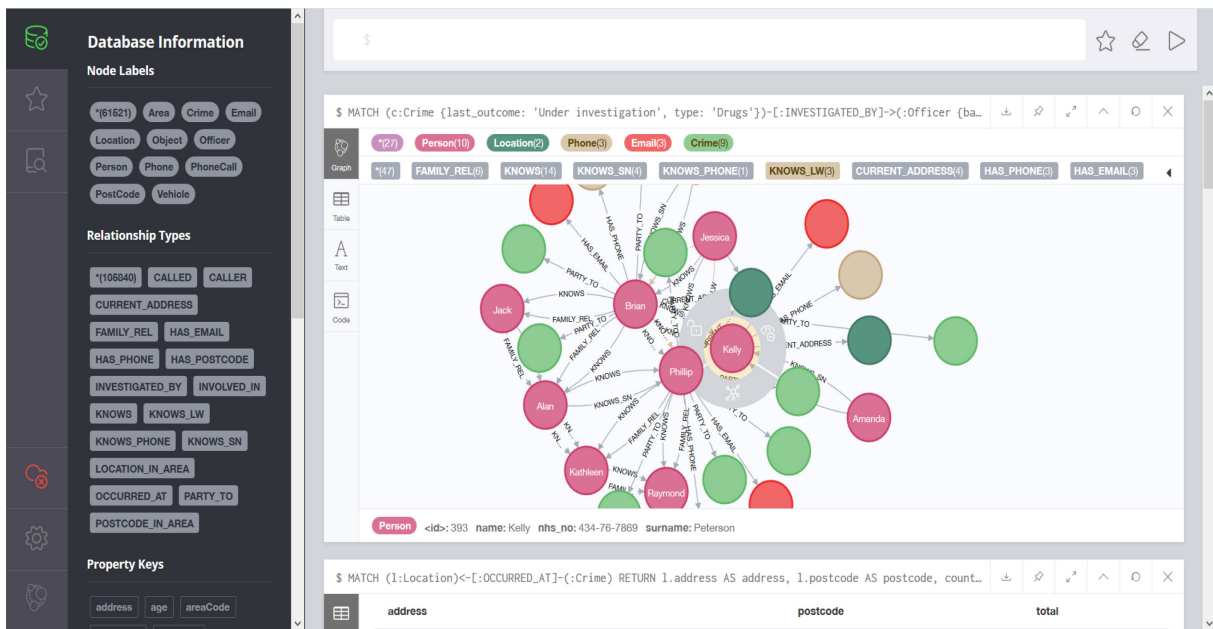


Abbildung 66: Neo4j

Technische Daten	
Art der Anwendung	Desktopanwendung (remote)
Aktivierung	Installation, Server Setup
Betriebssystem	Win, MacOS, Linux
Benötigte Anwendungen	
Lizenztyp	Open Source ()
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	begrenzt
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse
Einzelmethoden	netwA, GMod, geoA, tmpA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	Graphdatenbanken

Eingabe und Formate	
Eingabemodi	Script, Text, WIMP
Dateiformate	neo4j db
Import	txt, gexf, weitere
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	anspruchsvoll
Link	https://neo4j.com/

4.4.9. Nodegoat

Nodegoat ist ein webbasiertes, umfangreiches Tool zur räumlichen, chronologischen, relationalen Kontextualisierung von (Meta-)Daten, z.B. in Form geographischer, diachroner oder semantischer Netzwerkkarten. Grundlage ist ein strikt objektorientierter Ansatz, bei dem die Hierarchie der enthaltenen Daten durch entsprechende Strukturierung des Objektnetzwerkes festgelegt werden kann. Nodegoat ist frei als Webversion und als kostenpflichtige Serverversion über ein Service-Level-Agreement verfügbar.

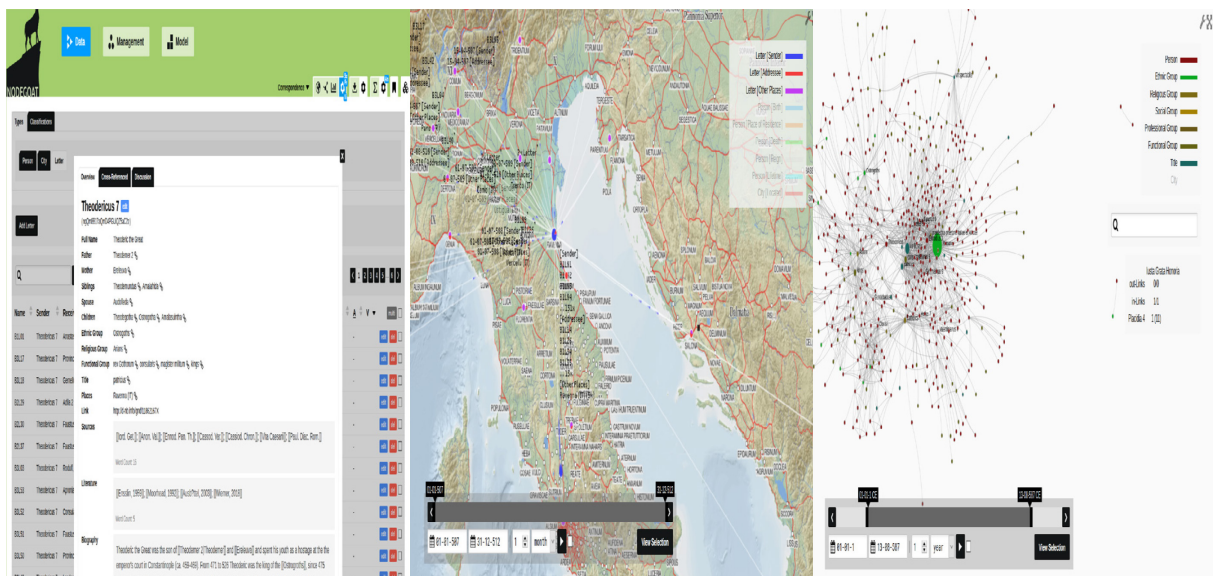


Abbildung 67: Nodegoat

Technische Daten	
Art der Anwendung	Website, Browseranwendung
Aktivierung	Registrierung, Server-Setup
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (GNU GPL v3.)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse, Annotation
Einzelmethoden	mAnn, sAnn, netwA, geoA,tmpA, Vis, Koll

Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP, Markup
Dateiformate	csv
Import	csv, odt
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	moderat
Link	http://nodegoat.net/

4.4.10. ORA-Lite Netscenes

ORA-Lite Netscenes ist ein Netzwerkanalysepaket zur Eingabe, Analyse und Visualisierung von Netzwerken mit Hilfe graphischer, statistischer und visualisierungsbasierter Techniken. Es erlaubt die Analyse multidimensionaler Metanetzwerke und ihrer Dynamiken unter Berücksichtigung zeitlicher und geographischer Veränderungen.

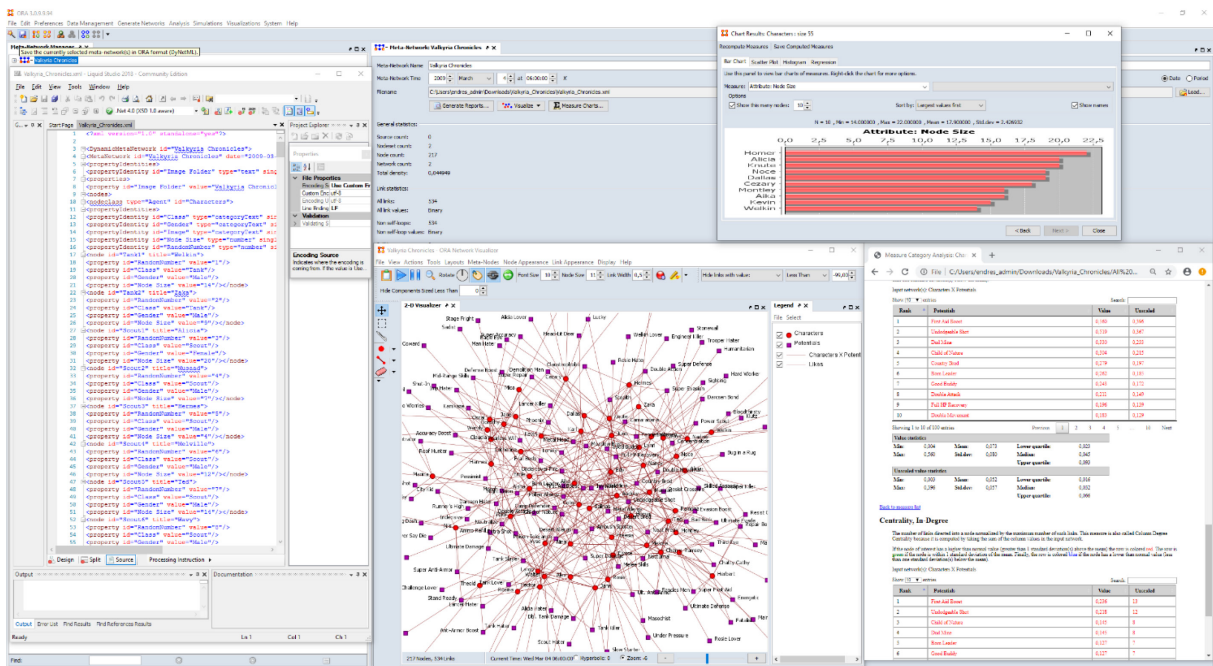


Abbildung 68: ORA Lite Netscenes

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	Win, Linux
Benötigte Anwendungen	Java
Lizenztyp	Closed Source, Freeware
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Netzwerkanalyse

Einzelmethoden	netwA, geoA, tmpA, GMod, Vis
Benötigte IT/DH-Vorkenntnisse	Graphdatenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	dl, xml
Import	xml, xls, dl
Export	xml, dl
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	komplex
Bedienung	schwer
Link	http://www.casos.cs.cmu.edu/projects/ora/software.php

4.4.11. Pajek

Pajek ist eine Open Source Software für Windows, MAC und Linux zur Analyse und Visualisierung bereits maschinenlesbar aufbereiteter Netzwerke, z.B. Text-, Personen-, Kollaborations-, Nachrichten-, Zitations- oder Diffusionsnetzwerke. Das Programm wurde seit der Veröffentlichung 1996 beständig weiterentwickelt. Hauptziel der Anwendung ist die Zerlegung umfangreicher Netzwerke in kleinere Einheiten und deren Weiterbearbeitung sowie die Bereitstellung leistungsstarker Visualisierungswerkzeuge und effizienter Algorithmen zur Analyse. Pajek erlaubt das Auffinden, Extrahieren und die visuelle Darstellung von Clustern. Das Programm unterstützt neben gerichteten und ungerichteten Netzwerken auch das Arbeiten mit skalierten bzw. bewerteten Netzwerken zwischen disjunkten Knotenmengen sowie die Analyse temporaler Netzwerke.

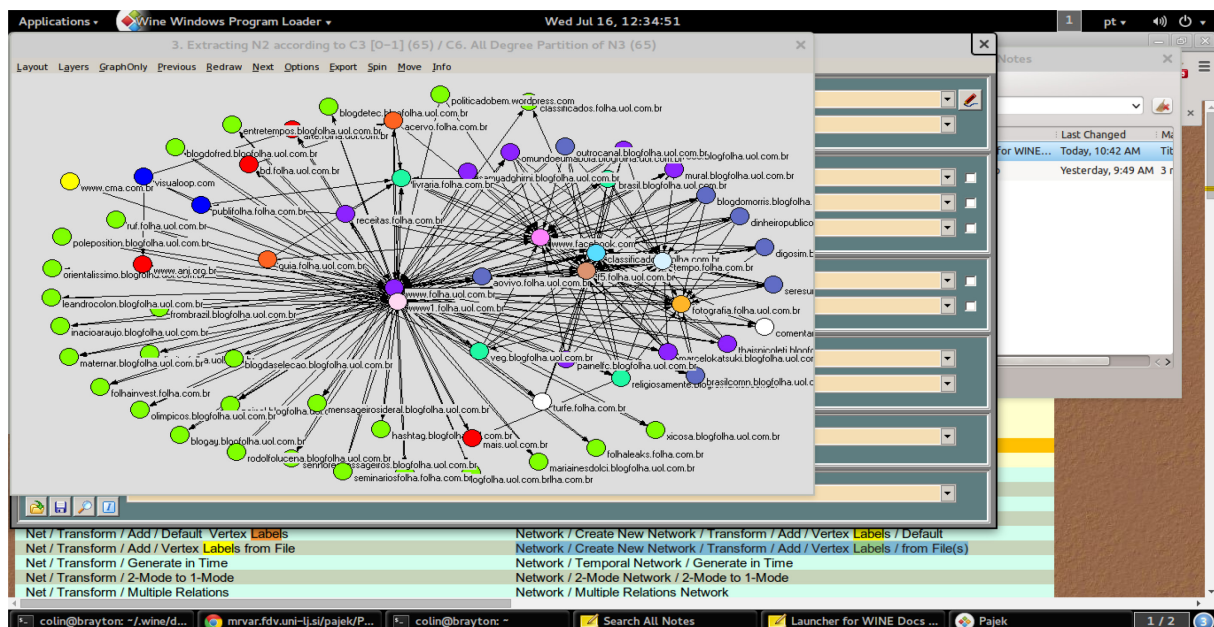


Abbildung 69: Pajek

Technische Daten	
Art der Anwendung Aktivierung Betriebssystem Benötigte Anwendungen Lizenztyp	Website Open Source (BSD 3)
Nutzerinnenunterstützung	
Dokumentation Integrierte Eingabe / Funktionshilfe Automatische Fehlerrückmeldung Support	elementar elementar nein nein
Zwecke und Nutzungsvoraussetzungen	
Methoden Einzelmethoden Benötigte IT/DH-Vorkenntnisse	Netzwerkanalyse, Geoanalyse netwA, geoA, tmpA, Vis
Eingabe und Formate	
Eingabemodi Dateiformate Import Export	Text, WIMP csv csv csv
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns Funktionsumfang Struktur der Steuerelemente Bedienung	domänenorientiert gering einfach leicht
Link	http://hdlab.stanford.edu/palladio/

4.5. Werkzeuge zur Georeferenzierung und Interaktive Digitale Kartenwerke

4.5.1. ArcGIS

ArcGIS ist ein geographisches Informationssystem (GIS), das die Arbeit mit Karten und geographischen Informationen ermöglicht und vom Environmental Systems Research Institute (Esri) verwaltet wird. Es kann für die Erstellung, Bearbeitung und Konfiguration digitaler Karten, datenbankbasierte Verwaltung von Geodaten, Visualisierung geographischer Informationen und zur Analyse aller Arten georeferenzierter Daten verwendet werden. Neben dem ArcReader zum Betrachten und Durchsuchen von Karten und verschiedenen ArcGis Desktop Versionen zur Erstellung von Karten und Layern, zur Datenverwaltung und -manipulation sowie zur Geodatenanalyse werden für den Serverbereich Lösungen zur gemeinsamen Arbeit in Datenbankmanagementsystemen angeboten. Aufgrund der vielfältigen Möglichkeiten zur Einbindung, Analyse und Visualisierung georeferenzierter Annotationen, Objekte und strukturierter Daten bietet sich die Anwendung für umfangreiche geschichts-, literatur- und sprachwissenschaftliche, archäologische, soziologische sowie kultur- und medienwissenschaftliche Forschungen an.

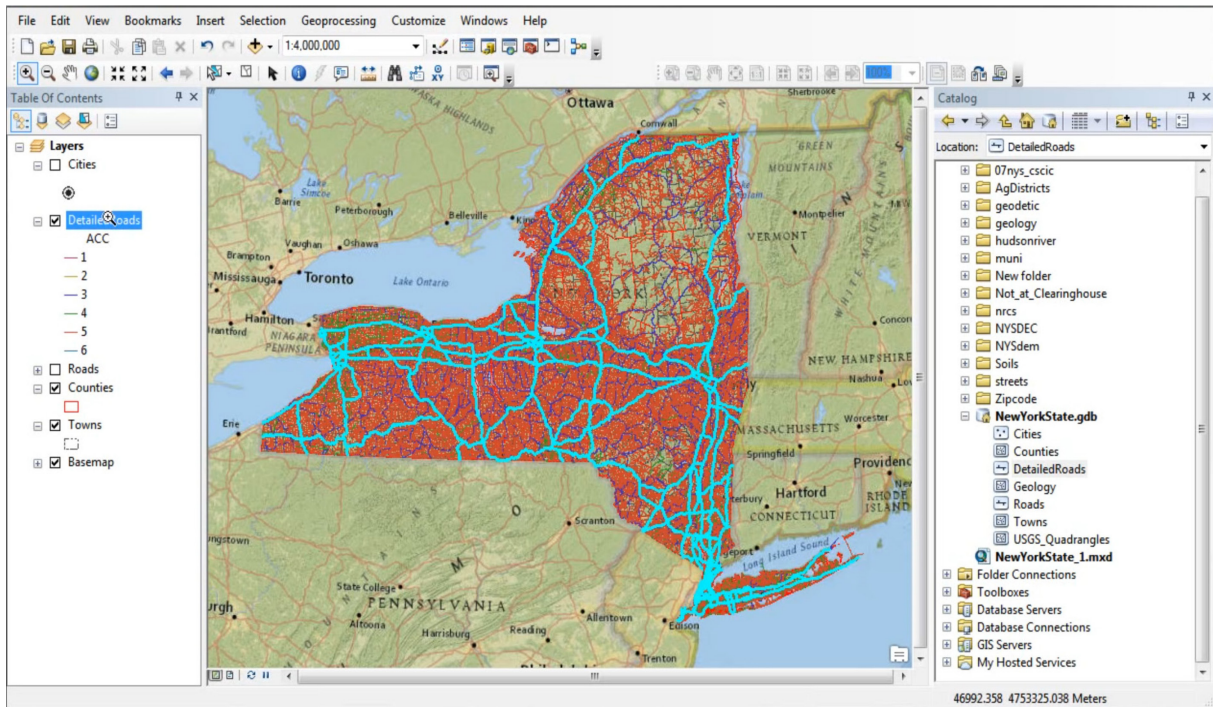


Abbildung 71: ArcGis

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal), Desktopanwendung (remote)
Aktivierung	Installation
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Closed Source, Payware
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, netwA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	Geodatenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	mxr, lyr, shp
Import	dbf, mdb, gdb, prn, shp, txt, csv, xls
Export	xml, zip, z, xls, tiff, img, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	anspruchsvoll
Link	https://www.esri.de/produkte/arcgis

4.5.2. DARE (Digital Atlas of the Roman Empire)

DARE ist eine präzise und dynamisch zoomfähige Karte Europas mit exakten Daten (Koordinaten, Zeitdaten, Typologie) zu Orten aus der Epoche des Römischen Imperiums und deckt den Zeitraum von 800 v.Chr. bis 700 n.Chr. ab. Neben den antiken Namen existierender und nicht mehr besiedelter Orte liefert es exakte Koordinaten, adressiert die Einträge in gängige Georeferenzierungssysteme und liefert Informationen zur Namenserwähnung historischer Orte.

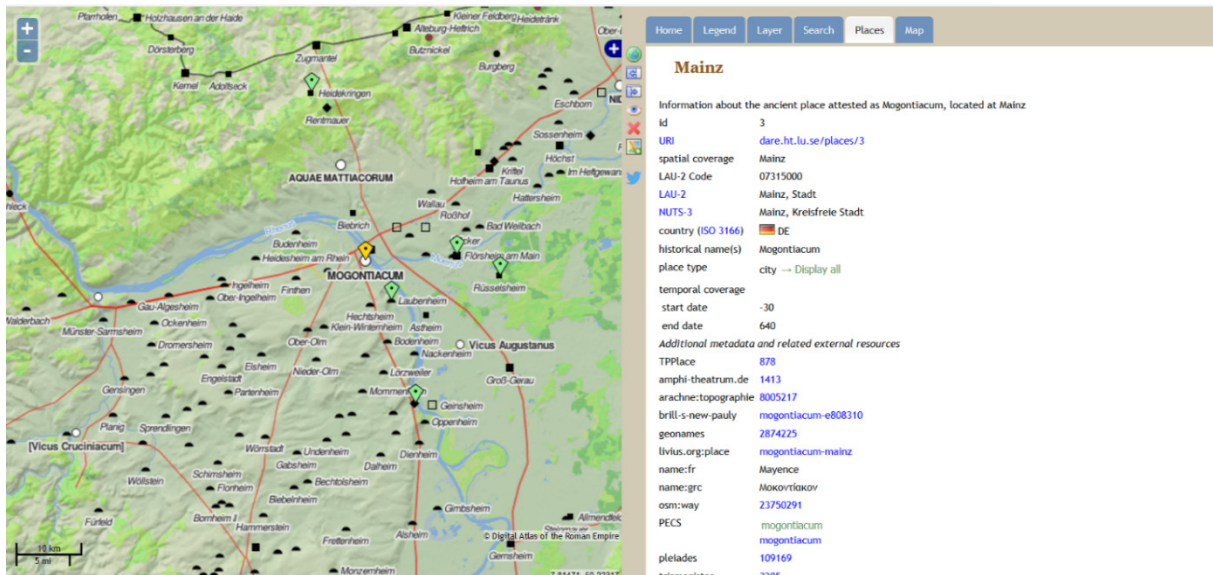


Abbildung 72: DARE (Digital Atlas of the Roman Empire)

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (CC BY 4.0)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	nein
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, Vis
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	WIMP
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	gering
Struktur der Steuerelemente	einfach
Bedienung	leicht
Link	https://dh.gu.se/dare/

4.5.3. DARIAH Geobrowser

Der DARIAH Geobrowser ist ein Werkzeug zur Analyse von Raum-Zeit-Beziehungen und erlaubt die räumliche und zeitliche Visualisierung von Forschungsdaten. Mit dem DARIAH-Datasheet-Editor können Daten eingegeben, importiert und weiterbearbeitet werden. Über eine Schnittstelle zum Getty Thesaurus of Geographic Names (TGN) ist die semiautomatische Georeferenzierung von Ortsnamen möglich. Auch die vergleichende Visualisierung mehrerer Forschungsdatensätze ist möglich, um Korrelationen zwischen den Daten im Hinblick auf ihre geographischen und zeitlichen Beziehungen darstellen und untersuchen zu können.

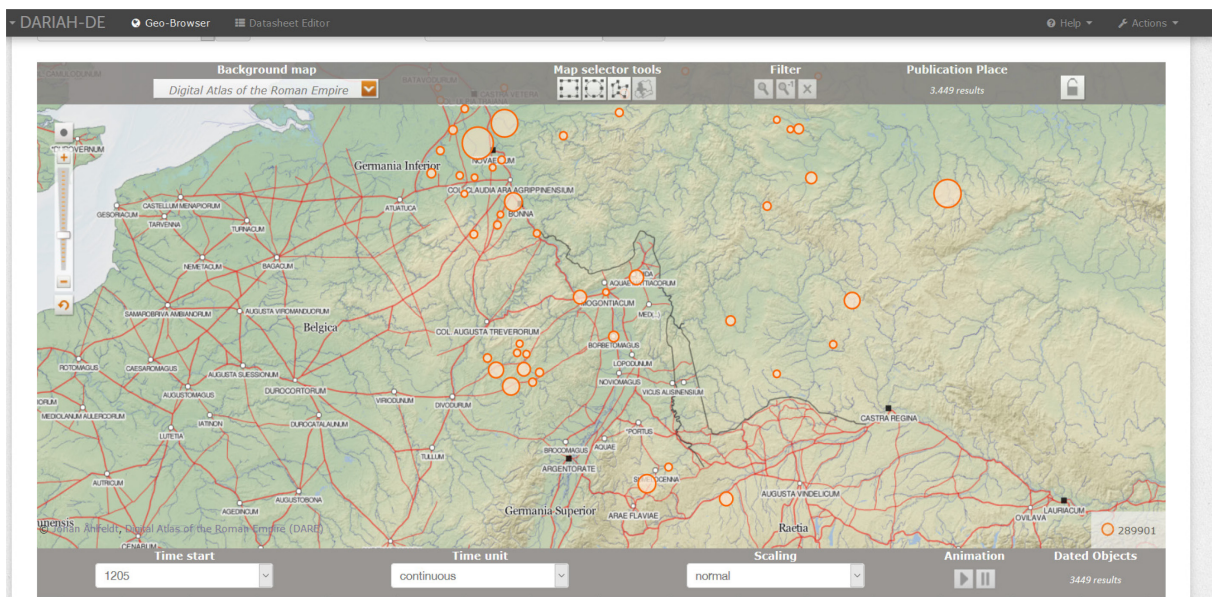


Abbildung 73: DARIAH Geobrowser

Technische Daten	
Art der Anwendung	Website
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	nein
Automatische Fehlerrückmeldung	nein
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, netwA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	svcasfa
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	kml
Import	kml, csv, xls
Export	kml, csv, xls
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	einfach
Bedienung	moderat
Link	https://geobrowser.de.dariah.eu/

4.5.4. Historypin

Historypin ist eine communitybasierte interaktive Webplattform zur Erstellung georeferenzierter Bildkollektionen. Über das Onlinetool können Objekte und Kommentare integriert und die Ortsbezüge der Inhalte kartiert werden. Die Anwendung bietet sich insbesondere für die Aufbereitung lokalhistorischer sowie literatur- und mediengeschichtlicher Themen an.

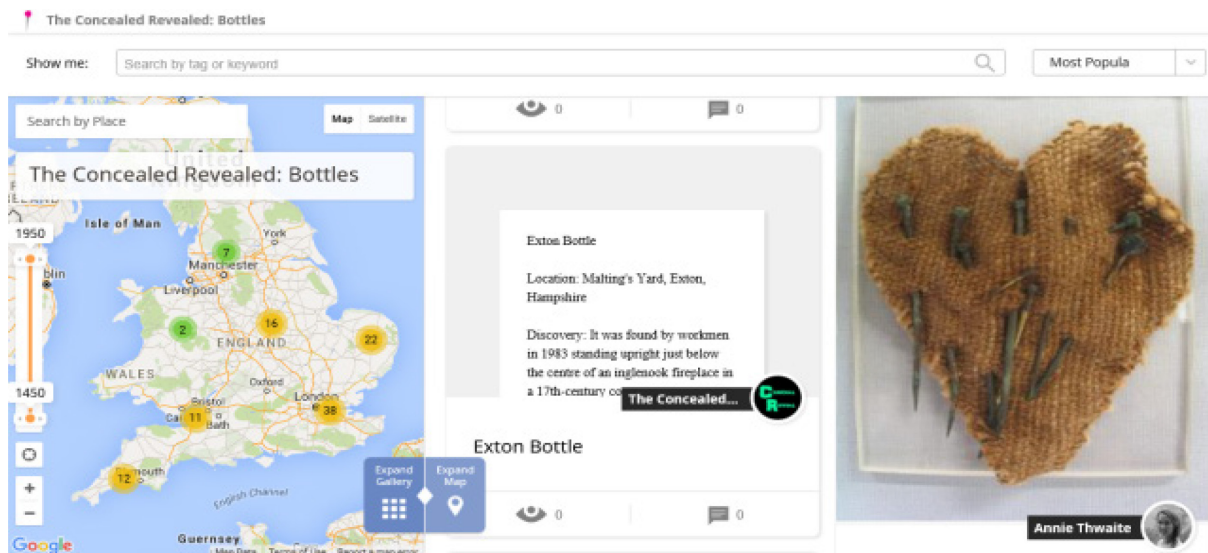


Abbildung 74: Historypin

Technische Daten	
Art der Anwendung	Website
Aktivierung	Registrierung
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (MIT)
Nutzerinnenunterstützung	
Dokumentation	elementar
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, tmpA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	moderat
Struktur der Steuerelemente	übersichtlich
Bedienung	leicht
Link	https://www.historypin.org/

4.5.5. Neatline

Neatline ist eine Anwendung zur Erstellung interaktiver Webausstellungen, mit der komplexe Karten, Zeitdaten, Artefakte, Bilder, Texte, Annotationen und Kommentare auf vielfältige Weise miteinander verknüpft werden können. Durch die mediale Kontextualisierung räumlicher, zeitlicher, graphischer und textueller Ebenen werden insbesondere Formen visueller Narration in der geisteswissenschaftlichen Forschung unterstützt. Damit eignet sich das Tool vor allem zur digitalen Aufbereitung von Kulturerbe aus alltags-, ereignis-, sozial-, politik-, personen- oder literaturgeschichtlicher Perspektive. Die Anwendung fungiert als Plug-In-Suite auf der Grundlage von Sammlungen und Archiven in der Publikationsplattform Omeka.

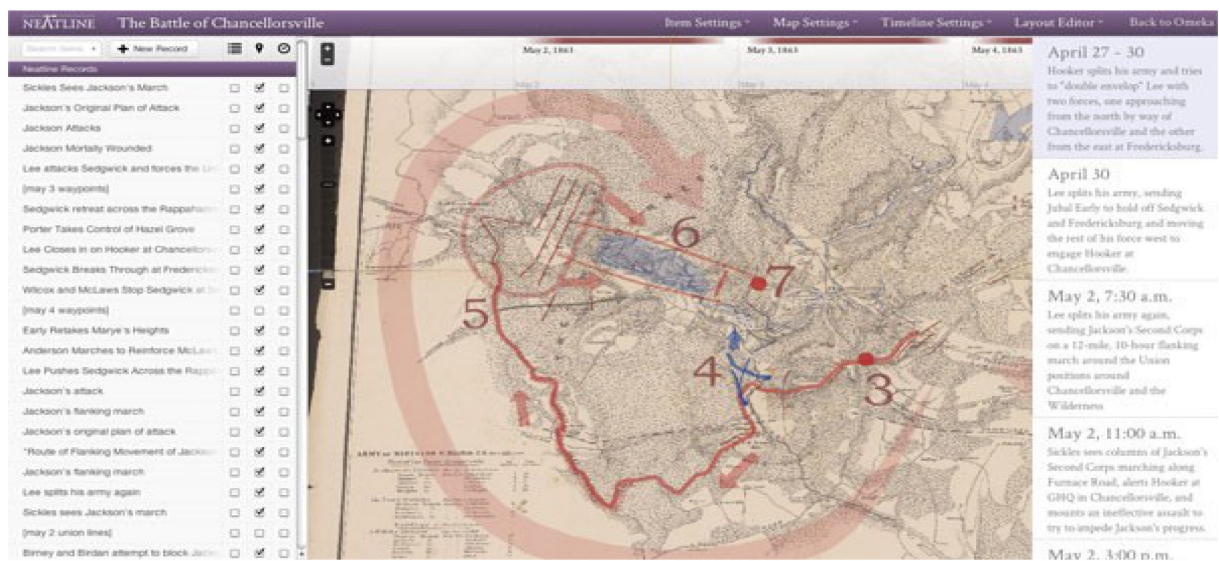


Abbildung 75: Neatline

Technische Daten	
Art der Anwendung	Desktopanwendung (remote)
Aktivierung	
Betriebssystem	
Benötigte Anwendungen	Omeka
Lizenztyp	Open Source (Apache 2.0)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	begrenzt
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, tmpA, Vis, Koll
Benötigte IT/DH-Vorkenntnisse	
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	
Import	
Export	
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	domänenorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	übersichtlich
Bedienung	moderat
Link	http://www.neatline.org/

4.5.6. QGIS

QGIS ist eine frei verfügbare Desktopanwendung für geographische Informationssysteme (GIS). Sie ermöglicht sowohl die Erfassung, Bearbeitung, Analyse und Visualisierung räumlicher Daten als auch die Einbettung, Erstellung und Transformation graphischer Karten oder anderer Hintergründe, auf denen Daten dargestellt werden. Das Programm unterstützt sowohl Raster- als auch Vektorlayer. Neben textbasierten Entitäten und Attributen sowie numerischen Werten können auch Bilder georeferenziert werden. Die Anwendung unterstützt zahlreiche Dateiformate und Webdienste, um Daten aus externen Quellen nutzbar zu machen. QGIS ist mit vielen anderen GIS-Paketen kompatibel und bietet je nach Bedarf zahlreiche Plug-Ins an.

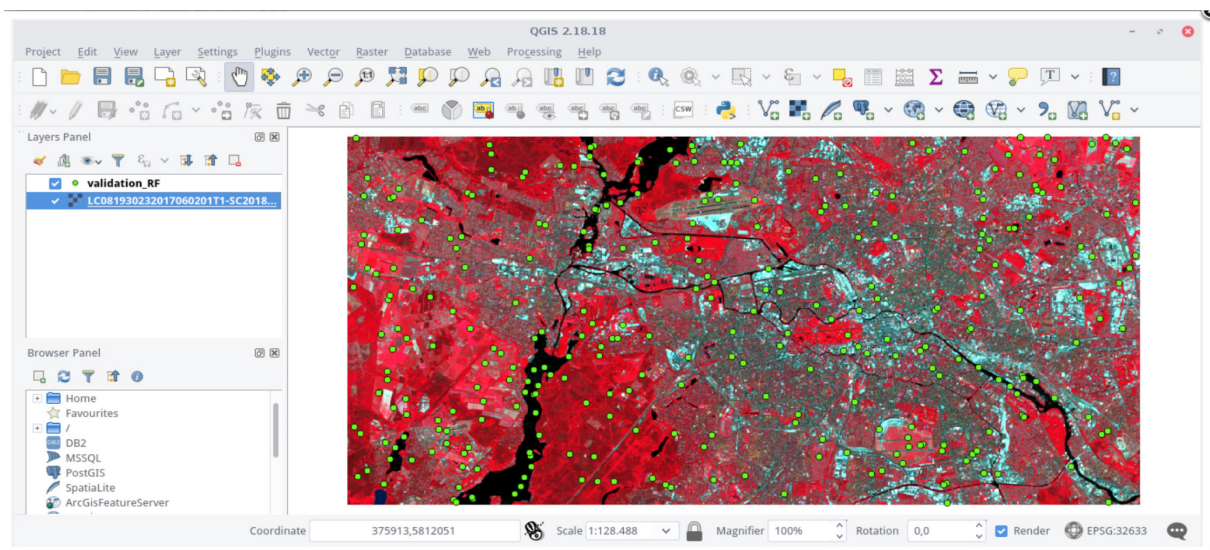


Abbildung 76: QGIS

Technische Daten	
Art der Anwendung	Desktopanwendung (lokal)
Aktivierung	Installation
Betriebssystem	
Benötigte Anwendungen	
Lizenztyp	Open Source (CC BY-SA 3.0)
Nutzerinnenunterstützung	
Dokumentation	ausführlich
Integrierte Eingabe / Funktionshilfe	elementar
Automatische Fehlerrückmeldung	ja
Support	elementar
Zwecke und Nutzungsvoraussetzungen	
Methoden	Geoanalyse
Einzelmethoden	geoA, netwA, tmpA, Vis
Benötigte IT/DH-Vorkenntnisse	Geodatenbanken
Eingabe und Formate	
Eingabemodi	Text, WIMP
Dateiformate	qgs
Import	shp, dbf, shx, ogr, txt, csv, weitere
Export	ps, png, svg, pdf, qgz, dxf, weitere
Konzeptionelle Gesichtspunkte der Anwendung	
Semantik des Informationsdesigns	funktionsorientiert
Funktionsumfang	hoch
Struktur der Steuerelemente	differenziert
Bedienung	anspruchsvoll
Link	https://qgis.org

5. Kurzauswertung der DH-Tool-Liste und Fazit

Abschließend soll anhand einfacher statistischer Auswertungen der in den Tabellen enthaltenen Informationen ein zusammenfassender Überblick zu den dargestellten digitalen Werkzeugen für die Geisteswissenschaften gegeben werden. Die nachstehende Grafik zeigt die Anteile der von den Anwendungen jeweils unterstützten geisteswissenschaftlichen Methoden. Anwendungen, die mehrere Methoden unterstützen, sind mehrfach berücksichtigt.

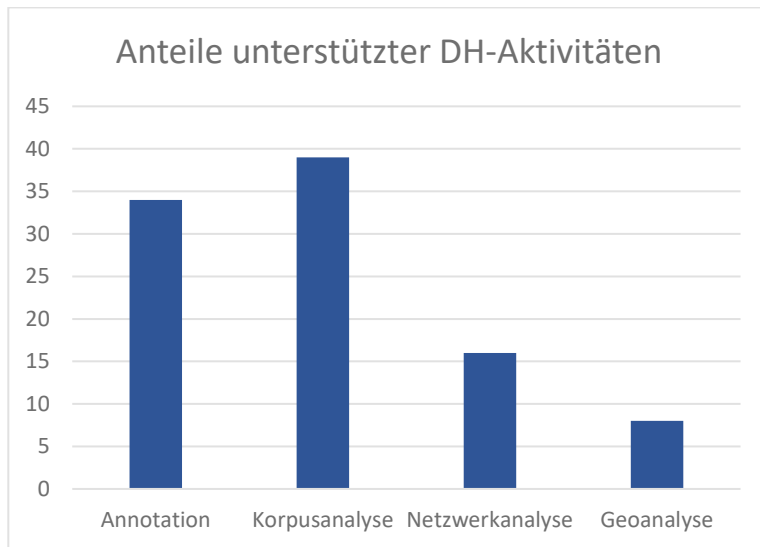


Abbildung 77: Anteile unterstützter DH-Aktivitäten

Die verschiedenen DH-Aktivitäten wurden im Rahmen der Beschreibung der Werkzeuge außerdem nach digitalen Einzelmethoden differenziert, deren Häufigkeiten im nachstehenden Diagramm dargestellt sind.

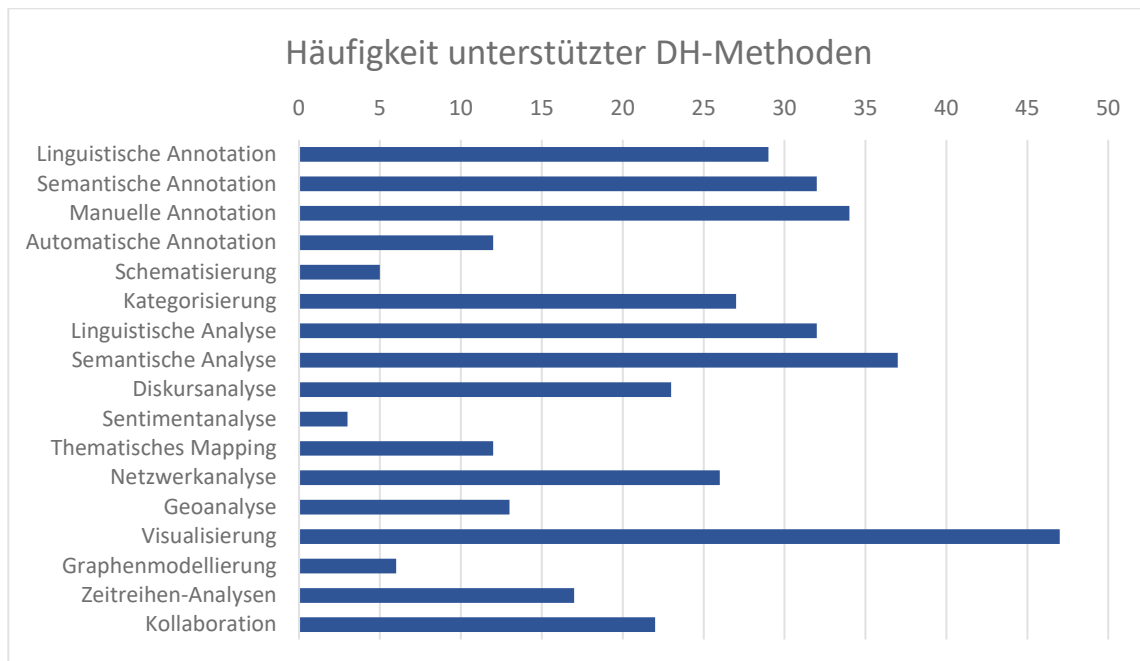


Abbildung 78: Häufigkeit unterstützter DH-Methoden

Auf der Ebene der Umsetzung dominieren vor allem lokale Desktopanwendungen und reine Webanwendungen, bei denen die Daten auf einem fremden Server verwaltet werden bzw. zur Archivierung vor dem Schließen des Browsers lokal gesichert werden müssen.

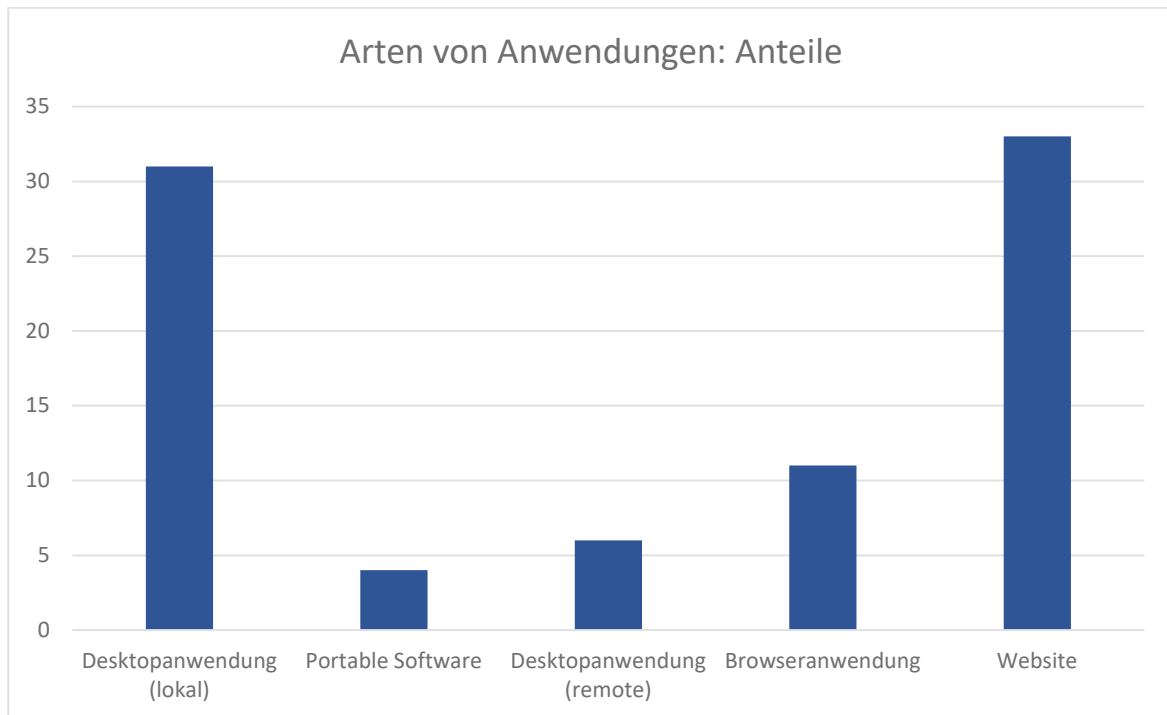


Abbildung 79: Anteile verschiedener Arten von Anwendungen

Dies schlägt sich auch in der Verteilung der Methoden zur Aktivierung der Anwendungen nieder: der überwiegende Teil der Anwendungen kann durch lokale Installation oder Registrierung beim Webdienst aktiviert werden, während direkt ausführbare Dateien und Skripte weniger stark vertreten sind.

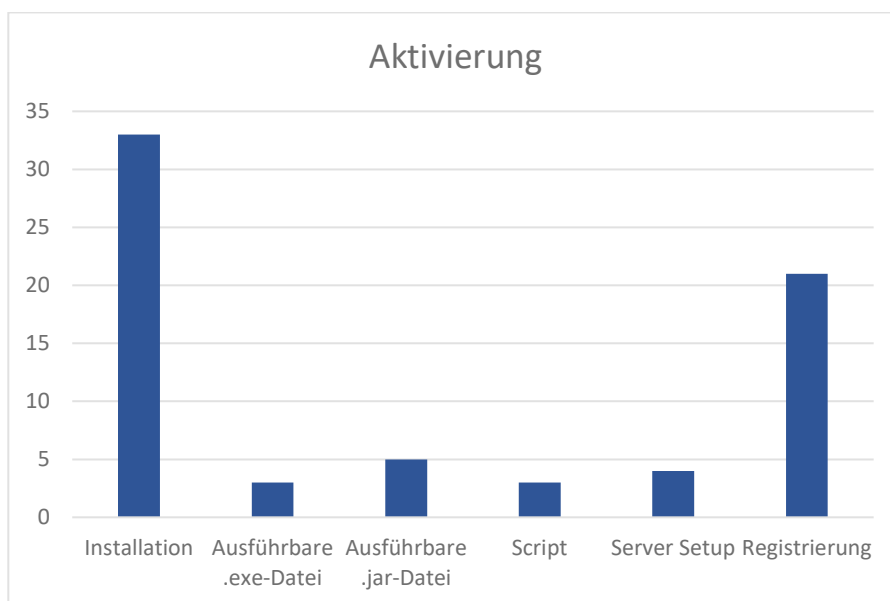


Abbildung 80: Anteile verschiedener Arten der Programmaktivierung

Bei den Arten der Dateneingabe ergibt sich ebenfalls ein relativ eindeutiges Bild: es dominieren die Reintexteingabe sowie die Eingabe über Steuerelemente (Windows, Icons, Menus und Pointer), während die Möglichkeit zur direkten Codierung von Text oder zur programminternen Erstellung eigener Scripts nur in wenigen Anwendungen vorgesehen ist.

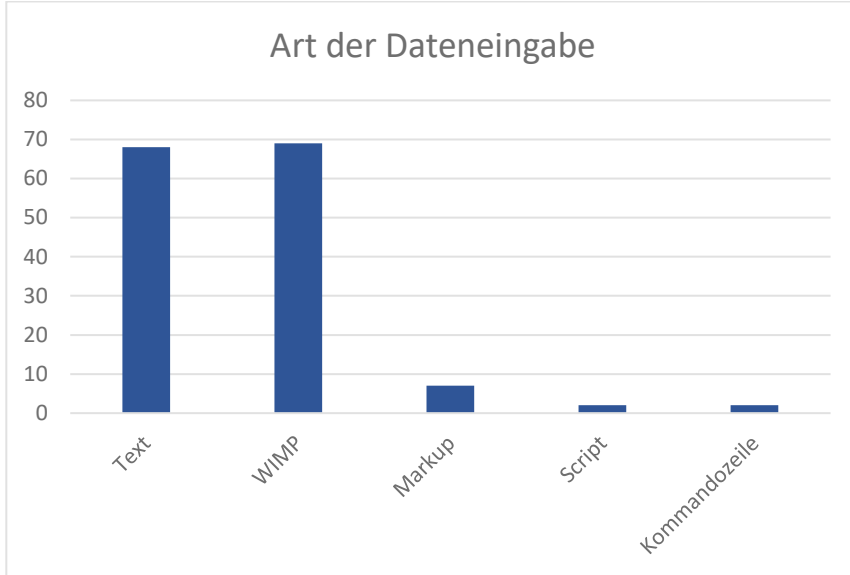


Abbildung 81: Unterstützte Arten der Dateneingabe

Die vorherrschenden Dateiformate sind xml, tei/xml, txt und csv. Nachfolgende Darstellung fasst die Häufigkeiten aller anwendungsintern sowie beim Im- und Export unterstützten Dateiformate zusammen. Deutlich wird der hohe Anteil weiterer, entweder programmeigener oder durch andere Anwendungen unterstützter Dateiformate.

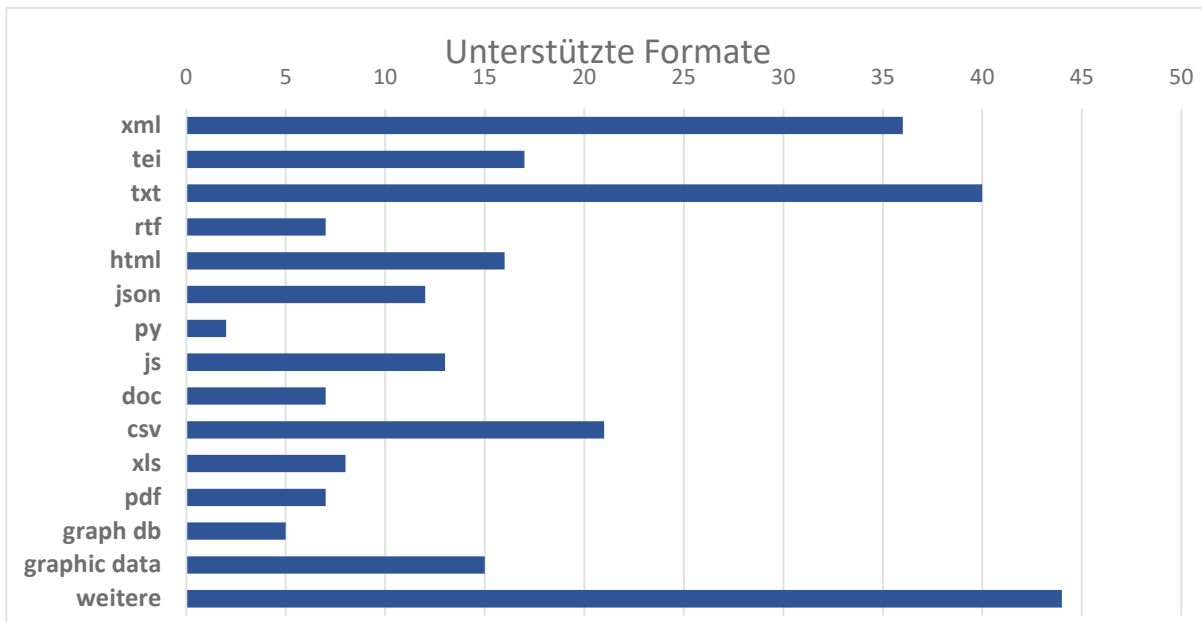


Abbildung 82: Häufigkeit der von den Anwendungen unterstützten Dateiformate

Für rund 90% aller evaluierten Anwendungen sind programmierungs- und bedienungsrelevante Dokumentationen vorhanden. Davon machen eher elementare Beschreibungen von Programmaufbau und Funktionalität etwa die Hälfte aus, während die andere Hälfte der Anwendungen sehr detailliert dokumentiert ist.

Zwei Drittel der Anwendungen enthalten integrierte Eingabeunterstützungen für die wesentlichen Funktionen, z.B. durch Dialoge in Kontextmenüs, Lightboxes oder Tooltips, aber auch integrierte Hilfemenüs. Bei einem Drittel der Werkzeuge sind solche unterstützenden Funktionen nicht oder nur in sehr begrenztem Maße vorhanden.

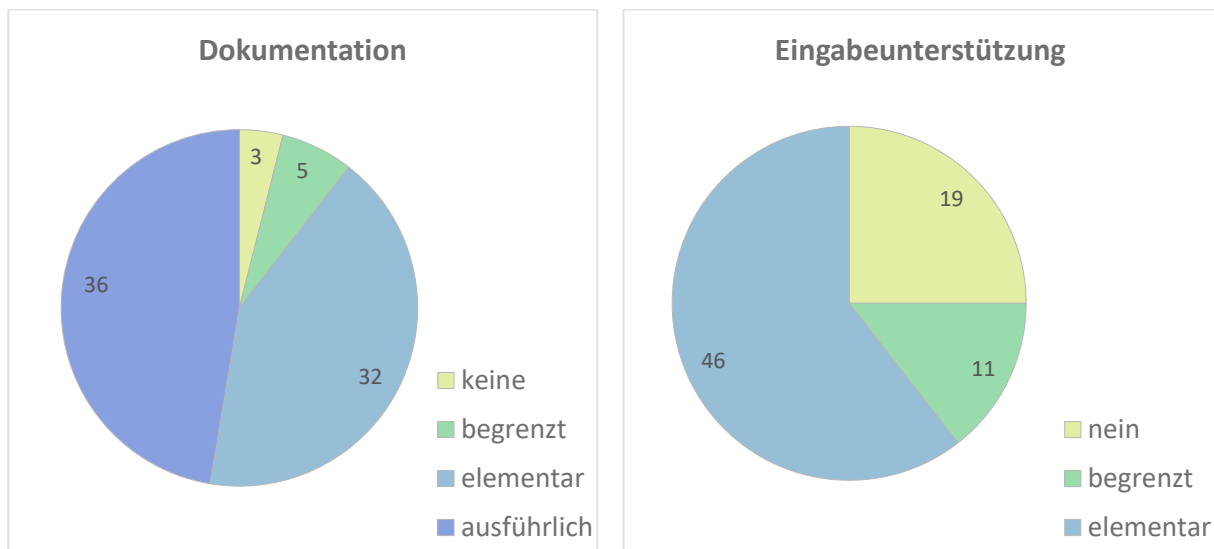


Abbildung 83: NutzerInnen-Unterstützung – Dokumentation und Eingabehilfe

Mehr als die Hälfte der Anwendungen bietet eine integrierte Fehlerkommunikation in Form von Pop-Up-Fenstern bzw. Infotabs, der Hervorhebung fehlerhafter Bereiche oder der automatisierten Kontextualisierung mit dem Hilfemenü an. Für knapp 60 % der Anwendungen stehen verschiedene Möglichkeiten des Supports zur Verfügung, wie z.B. durch Wikis, Knowledge Bases, FAQs, Videotutorials, grundlegende Hilfeseiten oder auch community-basierte NutzerInnenforen, in denen spezifische Funktions- und Editierungsprobleme behandelt werden. Für nur wenige der Tools sind Issue Tracking Systeme eingerichtet worden, über die bugs oder feature requests zwischen Anwendenden und Entwickelnden kommuniziert werden können. Je nach Umfang der zur Verfügung stehenden Unterstützungsmöglichkeiten ist der Support entsprechend entweder als begrenzt oder elementar eingestuft.

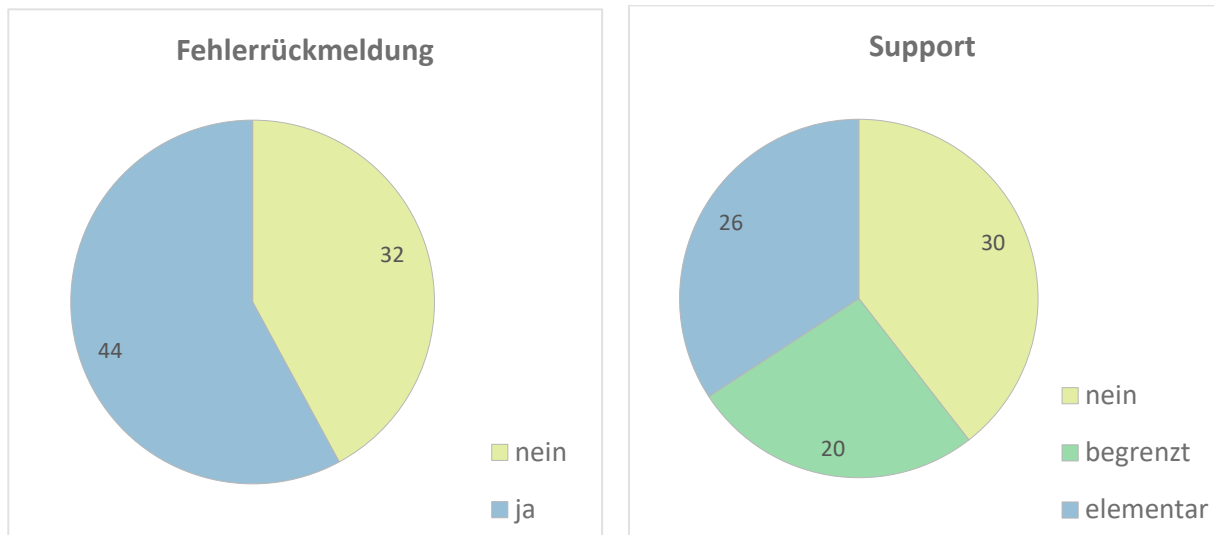


Abbildung 84: NutzerInnen-Unterstützung – Fehlerrückmeldung und Support

Weitergehende Formen des individuellen und umfassenden Supports sind in der Regel aufgrund des Fehlens von Langzeitressourcen nicht vorgesehen und stehen allenfalls in Verbindung mit kostenpflichtigen Service-Level-Agreements zur Verfügung.

Bei rund einem Drittel der Anwendungen werden Informatikkenntnisse vorausgesetzt. Diese Anforderungen beziehen sich entweder auf Kenntnisse von Programmiersprachen zur Aktivierung der Anwendung, z.B. bei der Ausführung spezieller Programmskripte oder Installationspakete, oder auf die Nutzung der Anwendung selbst, z.B. Kenntnisse von Markupssprachen, computerlinguistische Grundlagen oder auch Vorerfahrungen mit Graphdatenbanken und Datenmodellen. Der weitaus größere Teil kann jedoch ohne besondere Vorkenntnisse aktiviert und genutzt werden.

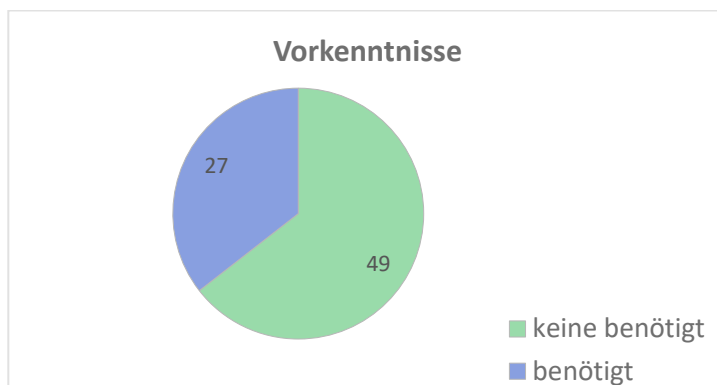


Abbildung 85: Bedarf an Vorkenntnissen zur Nutzung der Anwendung

Abschließend soll auf verschiedene konzeptionelle Gesichtspunkte der Anwendungen eingegangen werden. Ein wesentliches Differenzierungskriterium ist die Semantik des Informationsdesigns. Gemeint ist die fachsprachliche Ebene, über die Funktionen im Rahmen der Anwendung selbst sowie im Kontext der unterstützenden Formate wie Dokumentation und

Support vermittelt werden. Diese kann funktionsorientiert (fokussiert auf den technischen Prozess) oder eher domänenorientiert (fokussiert auf methodische Begriffe in den Geisteswissenschaften) organisiert sein. Das Spektrum innerhalb der Auswahl der Anwendungen reicht hierbei von eher an methodischen Prozeduren und Semantiken der Geisteswissenschaften orientierten Informationsdesigns bis hin zu sehr technisch gehaltenen sprachlichen Ausdrücken, die gewisse Kenntnisse der jeweiligen Funktionen voraussetzen. Die Zuordnung erfolgt danach, welche Tendenz in den jeweiligen Anwendungen stärker ausgeprägt ist.

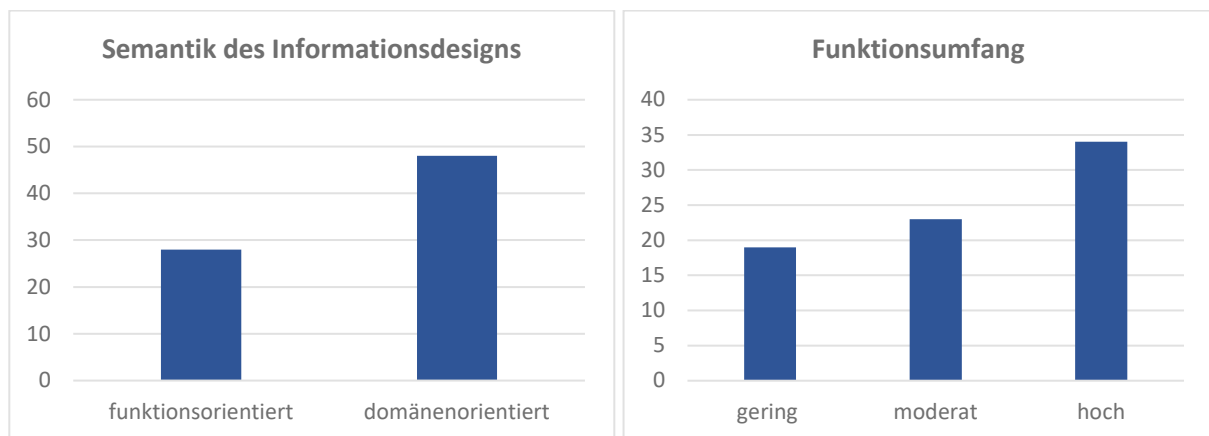


Abbildung 86: Semantik des Informationsdesigns und Funktionsumfang

Ein weiterer Gesichtspunkt ist der Umfang an Funktionen, der von den Anwendungen bereitgestellt wird. Dieser reicht von wenigen, in der Regel auf grundlegende methodische Konzepte beschränkte Funktionselemente bis hin zu sehr ausdifferenzierten Bearbeitungsmöglichkeiten und verschiedenen funktionalen Ebenen. Etwas mehr als die Hälfte der DH-Tools ist auf die digitale Realisierung grundlegender methodischer Schritte oder methodischer Bearbeitungsketten in den Geisteswissenschaften ausgerichtet und dementsprechend im Funktionsumfang gering bis moderat ausgeprägt. Ein nicht unwesentlicher Teil der Anwendungen bietet jedoch auch umfassende Funktionalitätsspektren an, die entweder durch methodische Breite mehrere geisteswissenschaftliche Erschließungsprozesse abbilden oder für einen spezifischen methodischen Bereich möglichst ausdifferenzierte Bearbeitungsmöglichkeiten anbieten.

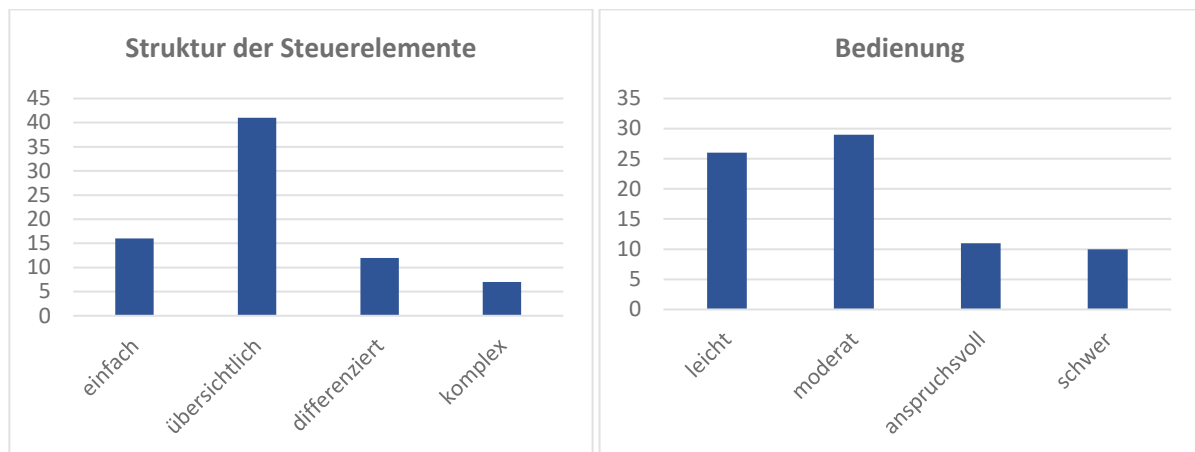


Abbildung 87: Struktur der Steuerelemente und Bedienung

Ein drittes Kriterium zur Einordnung der Anwendungskonzeption ist die Struktur der Steuerelemente der DH-Anwendungen. Diese ist zum Teil von der Semantik des Informationsdesigns als auch vom Funktionsumfang abhängig. Sie ergibt sich jedoch auch aus dem methodischen Kontext, der durch die Anwendung abgedeckt wird. Dementsprechend können Steuerungsstrukturen eher hierarchisch gegliedert oder durch ein Nebeneinander „gleichberechtigter“ Funktionsbereiche charakterisiert sein, was sich wiederum auf die Sichtbarkeit der Steuerungsebenen auswirkt. So können Nutzerschnittstellen mit vielen sichtbaren Steuerungselementen komplexer wirken als sie tatsächlich sind, während scheinbar einfach aufgebaute Interfaces mit zahlreichen Kontextmenüs und Subebenen hinterlegt sind. Auf Grundlage dieser Kriterien lassen sich etwa drei Viertel der Tools als einfach bzw. übersichtlich strukturiert bezeichnen. In diesen Fällen lassen sich die Steuerungsstrukturen relativ zügig überblicken und Funktionen den Bedienungselementen mehr oder weniger intuitiv zuordnen. Ein geringerer Teil der Anwendungen ist eher differenziert bzw. komplex gegliedert, so dass für die Erschließung der Steuerungsstruktur eine stärkere Zuhilfenahme von Dokumentation und Supportoptionen naheliegt.

Aus den drei vorgenannten Gesichtspunkten ergibt sich schließlich die Einschätzung der Bedienbarkeit der evaluierten DH-Anwendungen. Neben Informationsdesign, Funktionsumfang und Steuerungsstruktur spielen die vorhandenen anwendungsinternen und -externen Formen der NutzerInnenunterstützung wie Eingabehilfen, Fehlerrückmeldungen, Support und Dokumentation eine wichtige Rolle.

Eine ausführliche Dokumentation sowie schrittweise und an konkreten Daten geführte Anleitungen tragen somit indirekt zu einer erleichterten Bedienbarkeit auch komplex aufgebauter Werkzeuge bei. Umgekehrt kann jedoch die Bedienbarkeit von Anwendungen mit wenigen Funktionen und Steuerungselementen durch den Bedarf an Vorkenntnissen auf der Basis von Auszeichnungs- und Skriptsprachen durchaus anspruchsvoller ausfallen.

Die nachstehende Abbildung fasst das Verhältnis zwischen Bedienung und den anderen nutzungsrelevanten Gesichtspunkten der Anwendungen zusammen. Dabei wird deutlich, dass der Schwierigkeitsgrad der Bedienung bei Anwendungen mit funktionsorientierter Semantik, umfangreicherer Funktionalität und differenzierter Steuerungsstruktur zwar tendenziell höher ausfällt, allerdings kein eindeutiges lineares Verhältnis zwischen diesen Faktoren angenommen werden kann.

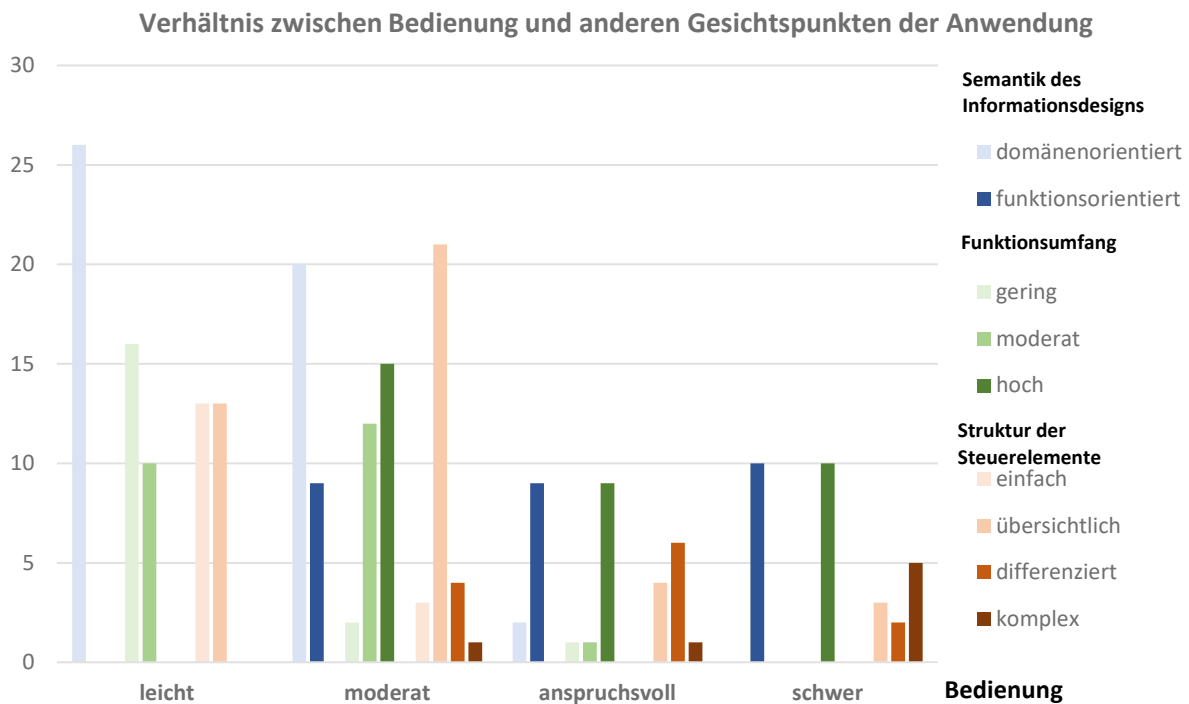


Abbildung 88: Zusammenhang anwendungsbezogener Gesichtspunkte

Abschließend sollen nochmals die Eigenschaften der vorliegenden Liste digitaler Tools und Ergebnisse ihrer Evaluation zusammengefasst werden. Die beschriebenen Werkzeuge bilden zwar wesentliche methodische Bereiche, aber zugleich nur einen Ausschnitt aus dem Gesamtprozess digitaler geisteswissenschaftlicher Arbeitsprozesse ab. Den Schwerpunkt bilden DH-Aktivitäten, die auch im Kontext des Projekts *Humanist Computer Interaction auf dem Prüfstand* im Fokus standen: Annotation, korpuslinguistische Analyse, Netzwerkanalyse und die Georeferenzierung von Textdaten. Digitale Werkzeuge für Methoden, die vor allem am Anfang und am Ende geisteswissenschaftlicher Arbeitsprozesse von Bedeutung sind, wie Transkriptionstools, Instrumente zur Recherche und Erfassung von Daten, Anwendungen zur Datenmodellierung und Plattformen zur digitalen Publikation und Archivierung stehen weniger im Vordergrund.

Ferner dürfte die Zusammenstellung einen leichten fachwissenschaftlichen Bias in Richtung von Methoden für die Geschichts-, Kultur-, Sprach- und Literaturwissenschaften aufweisen; diese

Tendenz ergibt sich ebenfalls aus dem Kontext des *Humanist*-Projektes und dem Bedarf der Aufbereitung und Erschließung einer frühmittelalterlichen Briefsammlung für die Entwicklung von Nutzerszenarien und die Validierung digitaler Methoden.

Darüber hinaus konzentriert sich die Auswahl auf Anwendungen, die zum größten Teil für GeisteswissenschaftlerInnen ohne fortgeschrittene Kenntnisse von Programmier- oder Markup-Sprachen zugänglich und aktivierbar sind. Der deutlich geringere Anteil an scriptbasierten Anwendungen (R, Python), Java-Applets oder command-line-basierten Tools ist jedoch auch dem Umstand geschuldet, dass solche Werkzeuge oftmals sehr projektspezifisch für bestimmte Einsatzbereiche entwickelt werden, anschließend trotz Open-Source-Lizenz für viele GeisteswissenschaftlerInnen wenig sichtbar sind und auch nicht ohne informationstechnische Unterstützung und entsprechende Anpassungen für eigene Projekte genutzt werden können.

Die Steckbriefliste stellt den Versuch dar, einen Querschnitt durch die DH-Tool-Landschaft zu geben, der verschiedene Aspekte und Qualitäten der digitalen Geisteswissenschaften hervorhebt. Sie verdeutlicht, dass es neben vielen quantitativen Analysewerkzeugen auch Ansätze für die in den Geisteswissenschaften besonders relevanten qualitativen Konzepte der Datenerschließung und Interpretation gibt. Sie veranschaulicht zugleich das äußerst breite Spektrum von sehr einfach gebauten und intuitiv ohne Weiteres bedienbaren Tools bis zu sehr komplexen und herausfordernden Instrumenten zur Erschließung und Analyse geisteswissenschaftlicher Daten. Ebenso legt sie die konzeptionelle Vielfalt digitaler Lösungsansätze sowie im Bereich der Strukturierung und Gestaltung von Nutzeroberflächen offen, durch die sowohl grundlegende Methoden der linguistischen und semantischen Textanalyse als auch spezifische Teilprobleme, Themenstellungen und methodische Zugänge innerhalb der Geisteswissenschaften bedient werden.

Nicht zuletzt bleibt diese Zusammenstellung aktuell verfügbarer DH-Werkzeuge lediglich ein Standbild zum gegenwärtigen Angebot in den digitalen Geisteswissenschaften. Ein im Kontext der Projektvernetzung immer wieder geäußertes Desiderat ist ein ständig aktualisiertes, detailliertes und durchsuchbares Archiv zu digitalen Tools, für das *DiRT Directory*, *TAPOR 3 Collection* und *Digital Toychest for the Humanities* wertvolle Beispiele sind, für die allerdings mittlerweile nicht mehr oder nur noch unregelmäßig Updates durchgeführt werden. Ein solches „Meta-Infrastruktur“-Projekt wäre entweder über eine institutionell verankerte Langzeitfinanzierung realisierbar oder als überinstitutionelle Onlineresource, die von der überwiegenden Mehrheit der EntwicklerInnen und NutzerInnen der DH-Community kollaborativ betreut und unterstützt würde.

6. Anhang

6.1. Literatur

Baum, C. & Stäcker, T. (2015). Die Digital Humanities im deutschsprachigen Raum. Methoden – Theorien – Projekte. In: C. Baum & T. Stäcker (Hrsg.), *Grenzen und Möglichkeiten der Digital Humanities (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1)*. DOI: [10.17175/sb001_023](https://doi.org/10.17175/sb001_023)

Bender, Michael (2020): Annotation als Methode der digitalen Diskurslinguistik. In: Gredel, Eva (Hg.), *Diskurse – digital, Bd. 2, Nr. 1, Diskurse - digital: Theorien, Methoden, Fallstudien*, S. 1-35. <https://majournals.bib.uni-mannheim.de/diskurse-digital/article/view/140>

Burghardt, M. (2014). Engineering Annotation Usability – Toward Usability Patterns for Linguistic Annotation Tools. Dissertationsschrift, Universität Regensburg. Online verfügbar unter <http://epub.uni-regensburg.de/30768/>

Gius, E. & Jacke, J. (2015). Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: C. Baum & T. Stäcker (Hrsg.), *Grenzen und Möglichkeiten der Digital Humanities (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1)*. DOI: [10.17175/sb001_006](https://doi.org/10.17175/sb001_006)

Gredel, Eva (2018). Digitale Methoden und Werkzeuge für Diskursanalysen am Beispiel der Wikipedia. In: M. Huber und S. Krämer, *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden. (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 3)*. DOI: [10.17175/sb003_005](https://doi.org/10.17175/sb003_005)

Kaden, Ben (2016). Zur Epistemologie digitaler Methoden in den Geisteswissenschaften. In: *Berliner Beiträge zu Digital Humanities (im Druck)*. Preprint: <http://dx.doi.org/10.5281/zenodo.50623>

Krämer, S. (2018). Der ‚Stachel des Digitalen‘ – ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Humanities in neun Thesen. *Digital Classics Online 4*. <https://doi.org/10.11588/dco.2017.0.48490>

Kuczera, A. et al. (Hrsg.). (2019). Die Modellierung des Zweifels - Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Zur Einführung. In: A. Kuczera et al. (Hrsg.). *Die Modellierung des Zweifels - Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4)*. DOI: [10.17175/sb004_013](https://doi.org/10.17175/sb004_013)

Nerbonne, John (2015). Die Informatik als Geisteswissenschaft. In: C. Baum & T. Stäcker (Hrsg.), *Grenzen und Möglichkeiten der Digital Humanities (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1)*. DOI: [10.17175/sb001_003](https://doi.org/10.17175/sb001_003)

Nielsen, Jakob (1995). „10 Usability Heuristics for User Interface Design“. Nielsen Norman Group. Website. <https://www.nngroup.com/articles/ten-usability-heuristics/>.

6.2. Abbildungsnachweise

Abbildung 1: eHumanities Desktop, Historical Semantics Corpus Management. Eigener Screenshot aus dem Humanist-Projekt

Abbildung 2: FUD Arbeitsumgebung. <https://fud.uni-trier.de/> (abgerufen am 20.04.2020)

Abbildung 3: GATE Developer GUI. Eigener Screenshot. <https://gate.ac.uk/family/developer.html>

Abbildung 4: Heurist Arbeitsumgebung. Eigener Screenshot. <http://heuristnetwork.org/>

Abbildung 5: TextGridLab. Eigener Screenshot.

Abbildung 6: TUSTEP Bearbeitungsbeispiel aus dem Berner Parzival Projekt, in: Viehhauser, G. (2013). Standardisierung und proprietäre Annotation im Berner Parzival-Projekt. Website.

<http://computerphilologie.digital-humanities.de/jg09/viehhauser.html> (abgerufen am 20.04.2020)

Abbildung 6: Annotate (A.nnotate). Eigener Screenshot. <https://www.annotate.com/>

Abbildung 7: Annotation Studio. Eigener Screenshot. <https://www.annotationstudio.org>

Abbildung 8: Annotations for Mac. Ausschnitt aus Demo-Video. <https://www.annotationsapp.com/>, <https://vimeo.com/103662588> (abgerufen am 20.04.2020)

Abbildung 9: Annotator.js Showcase. <http://annotatorjs.org/showcase.html> (abgerufen am 20.04.2020)

Abbildung 10: Brat Rapid Annotation, Ausschnitt aus Brat Demo.

http://weaver.nlplab.org/~brat/demo/latest/#/editable/CoNLL-ST_2002/ (abgerufen am 20.04.2020)

Abbildung 11: Catma Editor. Eigener Screenshot.

Abbildung 12: Classical Text Editor. <http://cte.oeaw.ac.at/> (abgerufen am 20.04.2020)

Abbildung 13: Coding Analysis Toolkit (CAT) Weboberfläche. Eigener Screenshot.

<https://cat.texifter.com/app/main.aspx>

Abbildung 14: Digital Mappa Publikationsbeispiel, <https://www.digitalmappa.org/> (abgerufen am 20.04.2020)

Abbildung 15: Ediarum Interface mit Bearbeitungsbeispiel. Dumont, S.; Fechner, M. (2015). ediarum. Eine digitale Arbeitsumgebung für Editionsprojekte. Vortragsreihe „Digital Humanities – Theorie und Methodik“, Leipzig, 20. Oktober 2016, Seite 5. Online Publikation. https://home.uni-leipzig.de/burr/Digital_Humanities/2015/ediarum_leipzig_2015-10-20.pdf (abgerufen am 20.04.2020)

Abbildung 16: eLaborate Bearbeitungsbeispiel. Dalen-Oskam, K. (Huygens Institute for the History of the Netherlands. Royal Dutch Academy of Arts and Sciences) (2012). Digital editions with eLaborate: from practice to theory. DHD 2012 Book of abstracts. <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/digital-editions-with-elaborate-from-practice-to-theory.1.html> (abgerufen am 20.04.2020)

Abbildung 17: Inception Demo Mode. Eigener Screenshot. <https://morbo.ukp.informatik.tu-darmstadt.de/login.html>

Abbildung 18: Morph Adorner Online Examples.

<http://morphadorner.northwestern.edu/morphadorner/onlineexamples/>

Abbildung 19: Natural Language Toolkit (NLTK). Beispiel aus Onlinedokumentation.

<https://www.nltk.org/> (abgerufen am 20.04.2020)

Abbildung 20: Nota Bene (NB). Ausschnitt aus "A guided tour to NB" (<http://nb.mit.edu/welcome>).

<https://vimeo.com/7370219> (abgerufen am 20.04.2020)

Abbildung 21: Now Comment Webplattform. Eigener Screenshot. <http://nb.mit.edu/welcome>

Abbildung 22: Pacte Weboberfläche. Eigener Screenshot. <https://patx-pacte.crim.ca>

Abbildung 23: Pundit Annotation Bearbeitungsbeispiel. <https://thepund.it/annotator-pro-web-annotation/> (abgerufen am 20.04.2020)

Abbildung 24: QAnnotate GUI. Eigener Screenshot.

- Abbildung 25: Recogito Weboberfläche. Eigener Screenshot. <https://recogito.pelagios.org/>
- Abbildung 26: Scripto Browser Oberfläche. Omeka S User Manual. <https://omeka.org/s/docs/user-manual/modules/scripto/scriptoproject/> (abgerufen am 20.04.2020)
- Abbildung 27: SPEEDy Standoff Properties Editor. Eigener Screenshot. <https://argimenes.github.io/standoff-properties-editor/>
- Abbildung 28: Text Analysis Markup System (TAMS) Analyzer. Project Sample. <https://sourceforge.net/projects/tamsys/> (abgerufen am 20.04.2020)
- Abbildung 29: Transkribus Bearbeitungsbeispiel. Transkribus Benutzeranleitung. <https://transkribus.eu/wikiDe/index.php/Benutzeranleitung#mediaviewer/Datei:Screenshot01.jpg> (abgerufen am 20.04.2020)
- Abbildung 30: UAM Corpus Tool GUI. Eigener Screenshot.
- Abbildung 31: Weblicht Analyse Umgebung. Eigener Screenshot. <https://weblicht.sfs.uni-tuebingen.de/weblicht/>
- Abbildung 32: WordFreak Interface. Example Screenshots. <http://wordfreak.sourceforge.net/screenshots.html> (abgerufen am 20.04.2020)
- Abbildung 33: ANNIS Frontend. Eigener Screenshot.
- Abbildung 34: Antconc GUI. Eigener Screenshot.
- Abbildung 35: CorpusExplorer. Screenshots der GUI. <https://notes.jan-oliver-ruediger.de/software/corpusexplorer-overview/> (abgerufen am 20.04.2020)
- Abbildung 36: CQP / Corpus Workbench Weboberfläche der TU Darmstadt. Eigener Screenshot
- Abbildung 37: Cosmas II (Corpus Search, Management and Analysis System) Weboberfläche. Eigener Screenshot.
- Abbildung 38: Hermetic Word Frequency Finder GUI. Eigener Screenshot.
- Abbildung 39: Jupyter Notebook. Linguistische Analysen des Humanist Projekts. Eigener Screenshot.
- Abbildung 40: JuxtaCommons Interface. Web-based Textual Analysis/Visualization Tools. <http://dayofdh2013.matrix.msu.edu/scholarscollab/tag/juxta/>
- Abbildung 41: LancsBox Dashboard. Eigener Screenshot.
- Abbildung 42: Rapid Miner Interface. Eigener Screenshot.
- Abbildung 43: RAW Graphs Weboberfläche. Eigener Screenshot. <https://app.rawgraphs.io/>
- Abbildung 44: Sketch Engine Web-Dashboard. Eigener Screenshot. <https://auth.sketchengine.eu>
- Abbildung 45: Stereoscope Weboberfläche und Analysebeispiel. http://www.stereoscope.threedh.net/index_de.html
- Abbildung 46: Stylo User Interface und Beispieloutput. <https://fortext.net/routinen/lerneinheiten/stilometrie-mit-stylo>
- Abbildung 47: Taporware Webinterface. Startpage, Interface und Output. Eigene Screenshots. <http://taporware.ualberta.ca/>
- Abbildung 48: Textable. Orange Interface und Visualisierungsbeispiel. <http://textable.io/>
- Abbildung 49: TextDNA. Eingabefeld und Visualisierung. https://graphics.cs.wisc.edu/Vis/SequenceSurveyor/TextDNA_GettingStarted.html
- Abbildung 50: TokenX Webtool. Eigener Screenshot. <http://tokenx.unl.edu/tokenx/index.html?file=../xml/base.xml>
- Abbildung 51: Tropes GUI und Visualisierungsbeispiel. Eigener Screenshot.
- Abbildung 52: TXM-Plattform. Exploration de corpus: outils et pratiques. Website. <http://explorationdecorpus.corpusecrits.huma-num.fr/txm/>
- Abbildung 53: Voyant Tools Weboberfläche. Eigener Screenshot. <https://voyant-tools.org/>
- Abbildung 54: Word and Phrase Web-Interphase. Eigener Screenshot. <https://www.wordandphrase.info/>
- Abbildung 55: WORDij 3.0 Interface. https://www.researchgate.net/figure/WORDij-30-Interface_fig1_248607689 . Visualisierungsbeispiel. <http://wordij.net/>
- Abbildung 56: Wordle Wordcloud. Eigener Screenshot. <http://www.wordle.net/>
-

- Abbildung 57: Wordtree Webtool. Eigener Screenshot. <https://www.jasondavies.com/wordtree/>
- Abbildung 58: Cytoscape Desktop GUI. Eigener Screenshot.
- Abbildung 59: EgoWeb 2.0 Interface. <https://www.youtube.com/watch?v=3hBgll3SbAw>
- Abbildung 60: Ezlinavis Webtool. Eigener Screenshot. <https://ezlinavis.dracor.org/>
- Abbildung 61: Gephi Destop GUI. Eigener Screenshot.
- Abbildung 62: GraphViz Editor. In: Russel, M.A.: Mining the Social Web. <https://www.oreilly.com/library/view/mining-the-social/9781449394752/ch01s02.html>
- Abbildung 63: Histogram Webtool. Eigener Screenshot. <http://histograph.eu/>
- Abbildung 64: InfraNodus Webtool. Eigener Screenshot. <https://infranodus.com/>
- Abbildung 65: Neo4j Sandbox. Eigener Screenshot. <https://sandbox.neo4j.com/>
- Abbildung 66: Nodegoat Web Interface und Visualisierungen. Eigene Screenshots. <https://nodegoat.net/>
- Abbildung 67: ORA-Lite Netscenes Desktop GUI. Eigener Screenshot.
- Abbildung 68: Pajek 3.1 Interface. Brayton, C. (2014): Pajek 3.1x. The Digital Structures of Media Organizations. Posted Content. <https://tupiwire.wordpress.com/2014/07/16/pajek-3-1x-getting-to-know-you/>
- Abbildung 69: Palladio Webtool. Eigener Screenshot. <https://hdlab.stanford.edu/palladio/>
- Abbildung 70: ArcGIS Beispiel. Basics 3a of 4. Cornell University. <https://www.youtube.com/watch?v=uWYY0h-9g4M>
- Abbildung 71: DARE Map. Eigener Screenshot. <https://dh.gu.se/dare/>
- Abbildung 72: DARIAH Geobrowser. Beispieldarstellung. Eigener Screenshot. <https://geobrowser.de.dariah.eu/>
- Abbildung 73: Historypin. Conceal Reveal. Historypin-Collection. <https://theconcealedrevealed.wordpress.com/historypin-collection/>
- Abbildung 74: Neatline Interface. <https://neatline.org/plugins/>
- Abbildung 75: Label Samples in QGIS. <https://blogs.fu-berlin.de/reseda/label-samples-in-qgis/>
-