

Usability-Analyse von digitalen Tools und Methoden in den Geisteswissenschaften

Ein Werkstattbericht

Tobias Simon¹, Jana Dolan¹, Alexander Schmitt¹ und Sven Pagel¹

Abstract: Fehlende Usability ist ein häufig auftretendes Problem von Tools aus dem geisteswissenschaftlicher Forschungskontext. In dieser Studie werden Tools, die im Editionsprozess eingesetzt werden, systematisch untersucht. Dabei wird auf einen Methodenmix aus Eyetracking und Gruppendiskussionen gesetzt. Die bisherigen Erkenntnisse geben einen Eindruck über typische Probleme, methodische Auswertungsmöglichkeiten und konkrete Empfehlungen für Verbesserungen der Tools.

Keywords: User Experience, Usability, Eyetracking, Digital Humanities

1 Usability in den Digital Humanities

Obwohl die Relevanz von Usability für die Akzeptanz digitaler Methoden in den Geisteswissenschaften mittlerweile oft betont wird [Bu16, Th17] und es eine Reihe von Untersuchungen zur Usability einzelner Anwendungen gibt, mangelt es nach wie vor einigen Tools an einer hohen Benutzerfreundlichkeit. Diese fehlende Berücksichtigung von Usability-Kriterien bei der Entwicklung von Anwendungen für die Digital Humanities (DH) wurde bereits häufig kritisiert (u.a. [Ed12, GO12, Th17]). Mangelnde Usability ist vor allem problematisch, weil dadurch die Zugänglichkeit der Anwendungen eingeschränkt wird [Ed12], was zu negativen Wirkungen auf die Reputation der Disziplin der Digital Humanities führen kann. Für Thoden u. a. muss ein Tool, das Usability-Kriterien entspricht, Forschende dabei unterstützen, ihre Ergebnisse entsprechend der verwendeten Methoden zu erzielen und dabei immer transparent über das Zustandekommen der Ergebnisse sein [Th17].

Eine zentrale Ursache für mangelnde Usability von DH-Anwendungen scheint der fehlende Austausch zwischen den Entwickler*innen von Software und potenziellen Anwender*innen zu sein. Schreibmann und Hanlon zeigen in ihrer Arbeit, dass nur ein sehr geringer Teil von Entwickler*innen systematisches Feedback zu den entwickelten Programmen erhebt. Lediglich die Hälfte der befragten Entwickler*innen von geisteswissenschaftlichen Tools betrachteten Toolakzeptanz und die Anzahl von Nutzer*innen als Erfolgsindikatoren. Ein Drittel gaben an, Usability-Analysen anzuwenden und lediglich

¹ Hochschule Mainz, Forschungsgruppe Wirtschaftsinformatik und Medienmanagement (WIMM), Lucy-Hillebrand-Straße 2, 55128 Mainz, {tobias.simon; jana.dolan; alexander.schmitt; sven.pagel}@hs-mainz.de

14% führten Befragungen von Nutzer*innen durch [SH10]. Dabei entstehen verschiedene Usability-Probleme: Burghardt unterscheidet zwischen allgemeinen und domänen-spezifischen Usability-Problemen [Bu12]. Während allgemeine Usability-Probleme durch das Einhalten generalistischer existierender Standards behoben werden können, ist für das Vermeiden von domänenspezifischen Usability-Problemen ein enger Austausch zwischen technischen und Domänenexpert*innen enorm wichtig. Einen ersten Ansatz für den Einbezug von Nutzer*innen liefert Bender [Be16]. In einer Bedarfserhebung, bei der 28 Geisteswissenschaftler*innen in halbstrukturierten Interviews befragt wurden, kann er sechs Bedürfniskategorien ableiten. Diese Kategorien sind Bedienbarkeit, Rezeption/Zugriff, Produktion/Motivation, Modularität/Anpassbarkeit, Kollaboration/Kooperation und Organisation/Verwaltung.

Die Forderung nach guter Usability hat unter anderem durch das enorme Angebot digitaler Anwendungen im Internet zugenommen und ist mittlerweile zu einer Grundvoraussetzung geworden [WX19]. Zur Messung der Usability von Anwendungen im DH-Kontext wurden bisher vor allem qualitative Studien durchgeführt (siehe z.B. [Bu16] für eine Übersicht über DARIAH-Studien zur Usability). Eine vielversprechende sowohl qualitative als auch quantitative Methode zur Usability-Erhebung ist daneben Eyetracking. Sie soll in der vorliegenden Studie eingesetzt werden, um neben qualitativen auch quantitative Aussagen über die Usability von ausgewählten DH-Anwendungen zu treffen. Dabei liegt der Fokus auf dem digitalen Editionsprozess, da dies im zugrundeliegenden Forschungsprojekt im Mittelpunkt des Erkenntnisinteresses steht.

Mithilfe quantitativer und qualitativer Methoden können wertvolle und tiefgründige Einblicke in die Usability der konkret untersuchten Software gewonnen werden. Allerdings ist ein softwareübergreifender Vergleich dabei oft schwierig. Das Ziel dieser Studie ist es daher, sowohl *qualitative als auch quantitative Aussagen zur Usability von Anwendungen für die Digital Humanities* zu treffen und daraus *Handlungsempfehlungen* zur Optimierung bestehender Tools sowie für die Entwicklung neuer Tools abzuleiten.

2 Studiendesign

2.1 Methoden

Im Rahmen der Studie sind insgesamt sechs nationale und internationale Workshops mit Geisteswissenschaftler*innen geplant, die jeweils unterschiedliche Vorkenntnisse (hoch, mittel, gering) im Umgang mit digitalen Methoden vorweisen. Die Workshops werden im Rahmen des interdisziplinären Projekts "Humanist Computer Interaction auf dem Prüfstand" durchgeführt, an dem die Johannes-Gutenberg-Universität Mainz, die TU Darmstadt sowie die Hochschule Mainz beteiligt sind. Dabei wird ein Methodenmix aus Schulung in digitalen Anwendungen, Befragungen, quantitativem und qualitativem Eyetracking sowie Gruppendiskussionen eingesetzt. Für die Bewertung der Usability sind vor allem die **Eyetracking-Messungen** (qualitativ und quantitativ) sowie die

Gruppendiskussionen (qualitativ) relevant. Zur Einordnung der Erhebungen dienen quantitative Fragebögen, mit deren Hilfe soziodemographische Faktoren und Vorkenntnisse erfragt werden. Dabei werden sowohl die Erfahrungen in Bezug auf digitale Methoden im Allgemeinen, als auch in Bezug auf die verschiedenen Anwendungskontexte ermittelt.

2.2 Untersuchte Anwendungen

In Anlehnung an Unsworth [Un00] und Stiller et al. [St15] wurden Anwendungen gewählt, die jeweils prägnante Aufgaben im geisteswissenschaftlichen Arbeitsprozess erfüllen sollen (siehe Abb. 1). Abgedeckt wurden die iterativen Phasen Annotieren, Kontextualisieren, Analysieren und Interpretieren. Tabelle 1 beschreibt die Zuordnung von Anwendungen und den entsprechenden Prozessschritten.

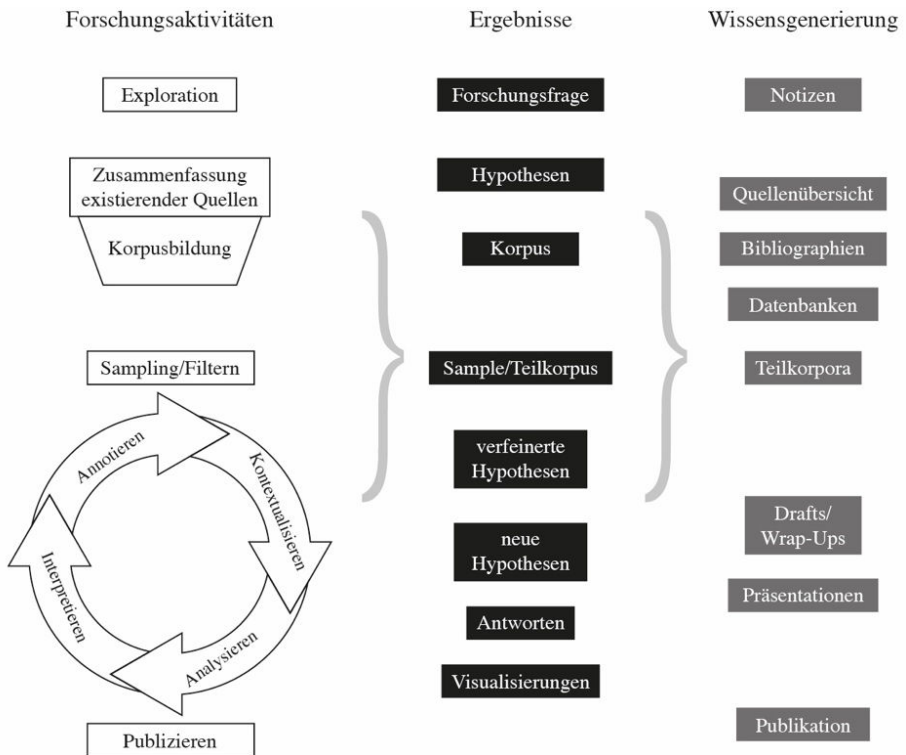


Abb. 1: Forschungsprozess in den Geisteswissenschaften [St15].

Anwendung	Prozessschritt
<i>Dillinger</i> ²	Annotieren
<i>QAnnotate</i> ³	Annotieren
<i>Nodegoat</i> ⁴	Kontextualisieren, Analysieren, Interpretieren
<i>Jupyter Notebook</i> ⁵	Kontextualisieren, Analysieren, Interpretieren

Tab. 1: Zuordnung von Anwendungen und Prozessschritten

Für die Wahl der Tools war vor allem eine leichte Zugänglichkeit entscheidend. Diese bezieht sich sowohl auf eine leichte Erlernbarkeit (bei gleichzeitig möglichst effektiver Erfüllung der Anforderungen) und Vermittelbarkeit im Rahmen der Workshops, da von sehr heterogenen technischen Vorkenntnissen ausgegangen werden musste, als auch auf den physischen Zugang, da die Teilnehmenden während der Workshops mit ihren eigenen Rechnern arbeiteten. *QAnnotate* wurde vor allem wegen der Erfahrungen innerhalb des Projektes und des Interesses an einer Auswertung des Tools ausgewählt.

Dillinger ist ein webbasierter HTML5 Markdown-Editor. Markdown ist eine textbasierte Markup-Sprache, die 2004 entwickelt wurde und von John Gruber als einfache Alternative zu HTML veröffentlicht wurde. Um den Einstieg in digitale Methoden zu erleichtern und ein generelles Verständnis für die digitale Auszeichnung von Text zu schaffen, wurde Markdown als einfache Annotationssprache gegenüber HTML oder XML gewählt.

Komplexer ist das im Rahmen der Studie bzw. des Forschungsprojekts entwickelte Tool „*QAnnotate*“, mit dem Texte über eine graphische Oberfläche annotiert werden können, während im Backend TEI-XML Dokumente erzeugt werden. Das Tool wurde entwickelt, um Anwender*innen den Einstieg in digitales Annotieren zu erleichtern, ohne vorher XML erlernen zu müssen.

Nodegoat ist eine webbasierte Datenmanagement-, Netzwerkanalyse- und Visualisierungsumgebung, die es Wissenschaftler*innen ermöglicht, Datensätze auf Basis ihres eigenen Datenmodells zu erstellen. Sie bietet relationale Analysemethoden mit räumlichen und chronologischen Formen der Kontextualisierung an. Durch die Kombination dieser Elemente in einer Umgebung sind die Wissenschaftler*innen in der Lage, komplexe Datensätze sofort relational und räumlich zu verarbeiten, zu analysieren und zu visualisieren. *Nodegoat* wurde ausgewählt, da es deutlich leichter zu bedienen ist als andere, mächtigere, aber komplexere Dienste.

Jupyter Notebook ist eine Open-Source-Webanwendung, mit der Dokumente, die Live-Code, Gleichungen, Visualisierungen und narrative Texte enthalten, erstellt und freige-

² <https://dillinger.io/>

³ Bisher nur als projektinterne Beta-Version verfügbar, eine Veröffentlichung wird angestrebt.

⁴ <https://nodegoat.net/>

⁵ <https://jupyter.org/>

geben werden können. Zu den Anwendungen gehören unter anderem: Datenbereinigung und -transformation, numerische Simulation, statistische Modellierung, Datenvisualisierung und maschinelles Lernen.

2.3 Teilnehmende

Bislang wurden drei (in Frankfurt am Main, Münster und Stuttgart) von insgesamt sechs geplanten Workshops (zusätzlich Wien, Gent, Edinburgh) durchgeführt. Dabei wurden die Daten von insgesamt 29 Proband*innen aus unterschiedlichen Forschungsfeldern erhoben. Der Großteil der Teilnehmenden beschäftigt sich mit historischen Forschungsthemen (siehe Abbildung 2). Die Proband*innen besitzen unterschiedliche akademische Grade, vom Bachelor-Abschluss bis hin zur Professur (siehe Abb. 3) mit 28 Prozent weiblichen Probandinnen und 72 Prozent männlichen Probanden und einem durchschnittlichen Alter von 37 Jahren.

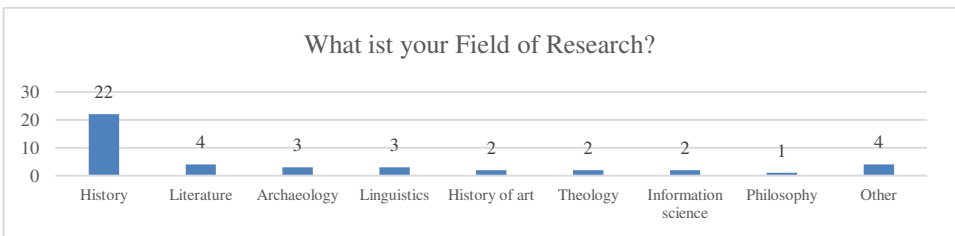


Abb. 2: Forschungsschwerpunkte der Proband*innen (n=29, Mehrfachantworten möglich)

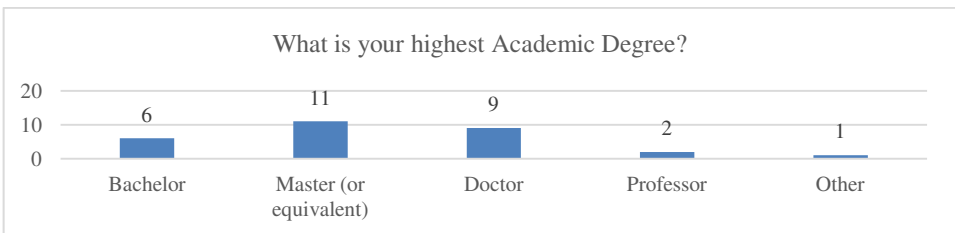


Abb. 3: Höchster Akademischer Grad der Proband*innen (n=29)

2.4 Eyetracking-Aufgaben

Die Aufgaben für die Eyetracking-Untersuchung wurden so gewählt, dass die Teilnehmenden der Workshops die zuvor in den Schulungsteilen erlangten Kenntnisse und Fähigkeiten über die Anwendungen gezielt einsetzen konnten. Dabei wurde besonders betont, dass es sich nicht um die Überprüfung der Proband*innen, sondern um eine Untersuchung der Anwendungen handelt. Folgende Aufgaben (siehe Tabelle 2) wurden

innerhalb des Workshops gestellt, wobei nicht alle in der Auswertung für diesen Beitrag berücksichtigt wurden (siehe Kap. 3).

Anwendung	Prozessschritt	#	Aufgabe
<i>Dillinger</i>	Datenimport	1.1	Bitte importieren Sie die Datei <i>variae-I-11.md</i> auf der Website <i>dillinger.io</i> über die Funktion „Import from Markdown File“.
	Auszeichnen	1.2	Zeichnen Sie die Überschrift des gegebenen Briefes als solche aus.
	Annotieren	1.3	Taggen Sie alle Personen, die im gegebenen Brief vorkommen, durch Kursivierung und alle Regionen durch Fettdruck.
	Annotieren	1.4	Erstellen Sie am Ende des Texts eine Fußnote, in der Sie alle genannten Personen auflisten.
<i>QAnnotate</i>	Annotieren	2.1	Taggen Sie alle Personen und Regionen, die in Brief I.11 vorkommen, mit der entsprechenden ID aus der jeweiligen Datenbank. Sollten Sie eine Person oder eine Region nicht in der Datenbank finden, legen Sie dafür bitte einen neuen Datenbankeintrag an.
<i>Nodegoat</i>	Analysieren	3.1	Ermitteln Sie das Verwandtschaftsverhältnis zwischen Theoderic the Great und Theodahad.
	Analysieren	3.2	Finden Sie heraus, an welchem Ort im Zeitraum von 509 bis 511 die meisten Briefe zum Thema („content“) „military and armed conflict“ gesendet wurden.

Tabelle 2: Übersicht der Eyetracking-Aufgaben (fett gedruckt sind die in diesem Beitrag vorgestellten Ergebnisse).

2.5 Datenerhebung

Beim Eyetracking wird die Bewegung der Augen aufgezeichnet, während die Proband*innen einen Stimulus betrachten. Ein großer Vorteil der Methode liegt in der Möglichkeit, menschliches Verhalten ohne die bewusste Reflexion bzw. Selbstauskunft der Testpersonen messen zu können [Sc03]. Im Rahmen der Studie werden zwei verschiedene Eyetracking-Verfahren eingesetzt: Zum einen mobiles Eyetracking mithilfe einer Brille, zum anderen stationäres Eyetracking mithilfe eines direkt am Monitor angebrachten Messgeräts.

Beim mobilen Eyetracking tragen die Proband*innen während der Erhebung eine mit Kameras ausgestattete Brille, sodass sowohl die Bewegungen der Pupille als auch das Blickfeld aufgezeichnet werden. In der vorliegenden Studie kam die Eyetracking Brille *Tobii Pro Glasses 2* zum Einsatz, um während der Schulungsteile u.a. das Verhältnis

zwischen Blick auf die Schulungsunterlagen und tatsächlicher Nutzung der Tools zu ermitteln. Für eine quantitative statistische Auswertung sind die mit diesem mobilen Eyetracker gewonnenen Daten beim vorliegenden Versuchsaufbau nicht geeignet, da die Fallzahl dafür zu gering ist. Im Kontext der Studie dienen die Aufzeichnungen vor allem der Messung, an welchen Stellen der Blick von der Anwendung vom Bildschirm abgewandt und ggf. auf weitere Hilfsmittel wie Zettel und Stift zurückgegriffen wird. Hier liegt der Fokus demnach eher auf einer qualitativen Analyse der gewonnenen Daten.

Nachdem die Proband*innen mit drei der vier vorgestellten Tools gearbeitet haben, wurde zusätzlich noch ein stationäres Eyetracking-System eingesetzt, bei dem die Usability der drei Tools anhand vordefinierter Aufgaben, die so in ähnlicher Weise auch in den Schulungsteilen vermittelt wurden, untersucht werden sollte. Dabei wird ein Gerät zur Aufzeichnung der Blickrichtung am unteren Bildschirmrand angebracht. Im Rahmen der Studie wurde der *Tobii Pro x2* mit einer Abtastfrequenz von 60 Hz eingesetzt. Die Augenbewegungen können so ohne unmittelbaren Kontakt zwischen Proband*innen und dem Gerät verfolgt werden. Dabei ist das nachvollziehbare Blickfeld auf die Fläche des Bildschirms beschränkt. Der Vorteil dieser Methode liegt in der Möglichkeit, Anwendungen mithilfe quantitativer statistischer Auswertungen zu vergleichen. Dabei wird in erster Linie zwischen Fixierungen (Momente, in denen die Augen ruhen) und Sakkaden (Momente zwischen zwei Fixierungen) unterschieden [EW07]. Durch die Kombination von mehreren Fixierungen und Sakkaden ergeben sich verschiedene Maße, die analysiert werden können, wie z.B. die Pfadanalyse, Fixationsdauer, Hotspots und Dichteverteilungen [Sc03]. Mittlerweile gibt es eine Reihe von Ansätzen, um Korrelationen zwischen Eyetracking-Messungen und Usability-Problemen herzustellen [EW07, PB06].

3 Ergebnisse

Die in diesem Beitrag vorgestellten Ergebnisse verstehen sich als Werkstattbericht, da zum Zeitpunkt des Verfassens erst drei von sechs Workshops durchgeführt wurden. Die Ergebnisse erheben daher keinen Anspruch auf Vollständigkeit und sollen einen Eindruck über die bis dato gewonnenen Erkenntnisse liefern. Aus diesem Grund wurden für die Usability-Analyse der Anwendungen im Bereich der Digital Humanities zunächst drei Aufgaben ausgewählt, die zum einen für den Editionsprozess besonders relevant sind und für die außerdem bereits erste Auswertungen durchgeführt wurden.

3.1 Ergebnisse der Eyetracking-Untersuchungen (qualitativ/quantitativ)

Im Rahmen der Eyetracking-Untersuchungen können sowohl qualitative (Blickpfade und Heatmaps) als auch quantitative (Dauer bis zum Auffinden verschiedener Funktionen) Erkenntnisse gewonnen werden. Für die Eyetracking-Untersuchungen konnten insgesamt 14 Proband*innen ausgewählt werden, deren Blickabtastrate (gaze sampling rate) bei 85% und mehr lag. Dadurch kann gewährleistet werden, dass ausreichend viele Blickpunkte erfasst wurden.

Aufgabe 1.1: Datenimport in *Dillinger*

Da es im Editionsprozess häufig vorkommt, dass Dokumente ausgetauscht werden, wurden die Proband*innen zunächst gebeten, ein vorbereitetes Dokument zur Auszeichnung und weiteren Bearbeitung in *Dillinger* zu importieren. Das Dokument war auf dem Desktop abgelegt und sollte über die Funktion „Import from... Markdown File“ eingelesen werden. Die Schaltfläche „Import from...“ befindet sich in der oberen rechten Ecke der Anwendung. In Abbildung 4 wird deutlich, dass viele Proband*innen die Funktion links oben im Hauptmenü vermutet haben.

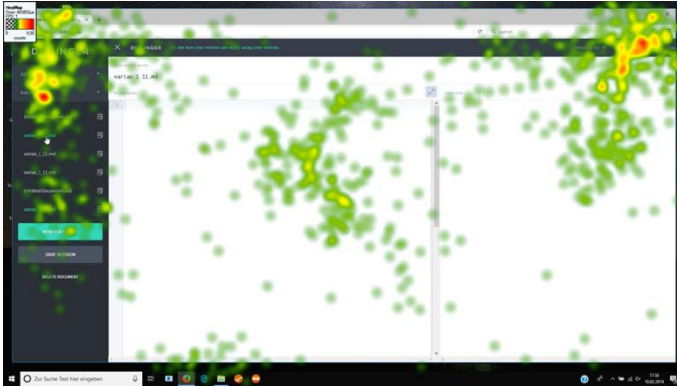


Abb. 4: Heatmap zum Auffinden der Import-Funktion in *Dillinger*

Problematisch dabei ist, dass die Import-Funktion rechts oben in den nicht-sichtbaren Bereich wandert, sobald das Hauptmenü auf der linken Seite geöffnet wird. Dadurch war es für die meisten Proband*innen schwer bis unmöglich, die Import-Funktion aufzufinden, wenn das Hauptmenü geöffnet war.

Aufgabe 1.3: Annotieren in *Dillinger*

Um zu überprüfen, wie schnell die Proband*innen die Personen und Regionen als solche auszeichnen konnten, wurden sogenannte Areas of Interest (AOI) gebildet. Die entsprechenden Wörter wurden als AOI gekennzeichnet, um die Zeit zu messen, bis die Proband*innen dieses Feld mit ihren Augen fixierten und anschließend zur Annotation angeklickt haben.

Tabelle 3 zeigt, dass die Proband*innen durchschnittlich 15,67 Sekunden zum Auffinden der entsprechenden Wörter im Text und durchschnittlich 26,55 Sekunden bis zum ersten Mausklick benötigten. Im Durchschnitt vergingen also 10,88 Sekunden von der ersten Fixation bis zum anschließend folgenden Annotieren. Betrachtet man sich die Daten genauer, wird jedoch deutlich, dass der Großteil der Proband*innen innerhalb kürzester Zeit (weniger als 4,31 Sekunden) den ersten Klick getätigt haben und es lediglich zwei

Ausreißer (Proband*in 2 und 7) gab, die nach der ersten Fixation länger benötigten, um die Wörter dann tatsächlich auszuzeichnen (siehe Abb. 5).

#	Proband*in	Time to First Fixation	Time to First Mouse Click	Differenz
1	101	7,34	10,90	3,56
2	105	7,47	90,81	83,34
3	108	61,53	62,5	0,97
4	112	5,57	8,72	3,13
5	122	12,19	14,56	2,37
6	123	6,18	8,98	2,80
7	125	38,18	76,82	38,64
8	128	11,2	12,59	1,39
9	133	15,41	16,26	0,85
10	134	8,04	9,94	1,90
11	136	10,61	14,08	3,47
12	138	20,65	22,06	1,41
13	139	6,45	10,63	4,18
14	140	8,52	12,83	4,31
	Mittelwert	15,67	26,55	10,88

Tab. 3: Zeit bis zur ersten Fixation und zum anschließenden ersten Klick zur Annotation in *Dillinger* (n=14).

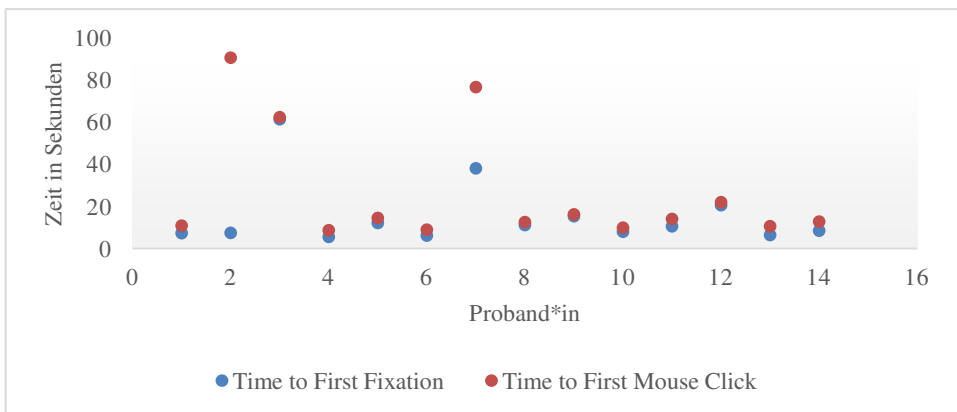


Abb. 5: Vergleich zwischen erster Fixation und erstem Klick in *Dillinger* (n=14).

Aufgabe 3.1: Analysieren in *Nodegoat*

Um das Verwandtschaftsverhältnis von *Theoderic the Great* und *Theodahad* zu ermitteln, war es notwendig über die Filterfunktion von *Nodegoat* die beiden Personen zu filtern. Im Durchschnitt benötigten die Proband*innen für das Auffinden des Buttons

16,11 Sekunden und für den ersten Mausklick 35,13 Sekunden. Dies deutet darauf hin, dass der Button zwar gefunden wurde, jedoch nicht mit der gewünschten Funktion in Verbindung gebracht wurde. Ein/e Proband*in fand den Button bereits mit sechs Blickpunkten nach 4,06 Sekunden und klickte auch unmittelbar nach 4,31 Sekunden auf die Filter-Funktion, im Gegensatz zu 27 Blickpunkten bei 23,76 bzw. 27,89 Sekunden (siehe Abbildung 6).

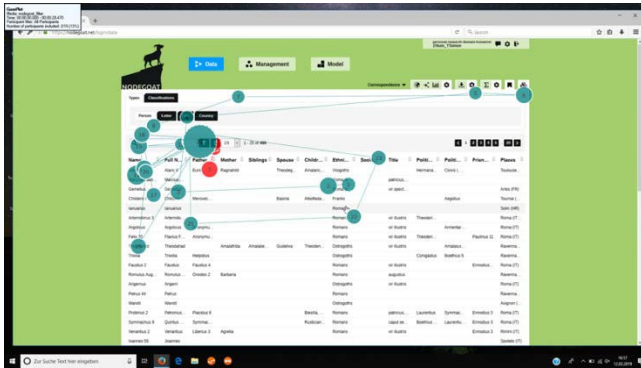


Abb. 6: Exemplarischer Gaze-Plot der Filterfunktion in *Nodegoat* mit sechs und 27 Blickpunkten.

Ein Blick auf die Heatmap (siehe Abbildung 7) zeigt, dass die meisten Proband*innen die Filterfunktion an der richtigen Stelle der Anwendung gesucht haben, es aber auch einige Ausreißer zur Navigationsebene rechts oben gegeben hat. Einige Proband*innen vermuteten die Möglichkeit nach Personen zu filtern jedoch auch in der Tabellenansicht, um über die Sortierung das Verwandtschaftsverhältnis zu ermitteln, was darauf schließen lässt, dass die Funktion nicht für jeden Nutzertypen unmittelbar verständlich ist.

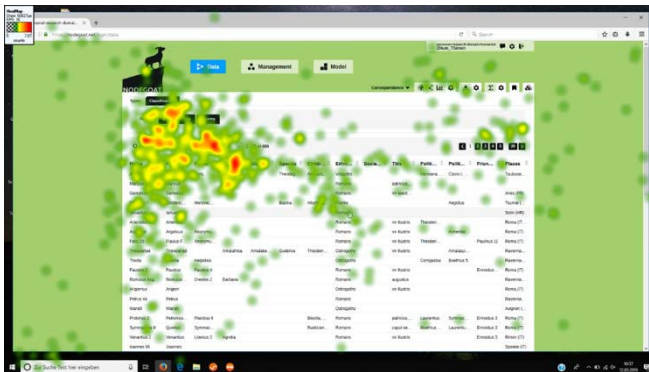


Abb. 7: Heatmap zum Auffinden der Filter-Funktion in *Nodegoat*

3.2 Ergebnisse der Gruppendiskussion (qualitativ)

Die Gruppendiskussion wurden jeweils nach den praktischen Anwendungsteilen mit allen Teilnehmenden (8-15 Personen) geführt und mithilfe eines Leitfadens moderiert.

Aus den Ergebnissen der Gruppendiskussion lassen sich, im Gegensatz zum Eyetracking, bewusste Reflexionen und Einstellungen der Workshopteilnehmenden gegenüber den verwendeten Tools ableiten. Dabei wurde sowohl gezielt über die einzelnen, im Workshop angewandten, als auch allgemein über die Erfahrung mit anderen digitalen Methoden gesprochen. Die Aussagen wurden durch qualitatives Clustering eingeordnet. Neben der Sortierung der besprochenen Anwendungen können so drei Kategorien gebildet werden: Verhältnis zwischen Aufwand und Nutzen, Anpassbarkeit und Interoperabilität.

Verhältnis zwischen Aufwand und Nutzen

Vor dem Hintergrund, dass im Laufe des geisteswissenschaftlichen Arbeitsprozesses (siehe Abb. 1) eine Vielzahl verschiedener Methoden und entsprechender Werkzeuge zum Einsatz kommt, ist die Bereitschaft, sich lange in komplexe Programme einzuarbeiten, mit denen jeweils nur einzelne Aufgaben bewältigt werden können, begrenzt. Dieses Problem wurde vor allem bei den Python-Skripten mit *Jupyter Notebook* angesprochen. Zwar wurde die Möglichkeit einer textübergreifenden Stilanalyse sehr positiv bewertet und der mögliche Mehrwert für die Forschung betont, allerdings wurde die aus Python-Code bestehende Benutzeroberfläche als abschreckend wahrgenommen. So äußerte eine Teilnehmerin folgendes:

„Ich denke man kann zielgerichtet sehr differenzierte Analysen durchführen und diese dann auch visualisieren. Also ich find das schon sehr nützlich aber [...] der Arbeitsanteil, den ich reinstecken muss, um Ergebnisse zu bekommen, den muss ich für mich persönlich abwägen, ob das meiner Fragestellung entspricht.“ (Diskussion M02, 00:08:07)

Ähnlich wurde die Situation von einem anderen Teilnehmer bewertet:

„[...] wenn ich solche linguistischen Analysen durchführen wollte, dann ist das meistens ein kleiner Teil meiner Arbeit. Und dafür dann den Aufwand zu betreiben, eine Programmiersprache zu lernen oder einen Informatiker zu fragen, mir ein Programm zu schreiben, das lohnt sich dann wahrscheinlich eben nicht.“ (Diskussion F03, 00:05:08).

Die grafische Oberfläche von *Nodegoat* wurde als deutlich ansprechender empfunden. Allerdings wurde eine teilweise irreführende Menüführung beim Hinzufügen neuer Parameter und die unübersichtliche Visualisierung von Abfragen kritisiert, weil deutlich mehr Optionen als benötigt angezeigt werden. Davon zeugen die beiden folgenden Aussagen:

„Am Anfang suggeriert es [Nodegoat], es sei total einfach. Sobald man dann aber den zweiten oder dritten Parameter ändern wollte, war es gleich kompliziert.“ (Diskussion M01, 00:11:20)

„[...] ich fand teilweise die grafische Darstellung schon sehr unübersichtlich, weil mehr Informationen angezeigt wurden als eigentlich gefragt wurden und das ging dann schon ein bisschen zulasten der Erkenntnis.“ (Diskussion M01, 00:29:40)

Auch hier wurde der wissenschaftliche Mehrwert im Verhältnis zu dem Aufwand, sich einen Überblick über die Funktionsweise des Programms zu verschaffen, kritisch bewertet. Bei *QAnnotate* hingegen wurde die Usability vor dem Hintergrund des Nutzens größtenteils positiv bewertet.

Anpassbarkeit

Als einer der Gründe für die positive Bewertung von *QAnnotate* wurde die übersichtliche und für Projekte individuell anpassbare Oberfläche angeführt.

„Ich fand QAnnotate ein ziemlich beeindruckendes Tool, weil man da sehr schön sehen konnte, dass das aus der Forschung gewachsen ist. Also man hatte irgendwie das Gefühl, da haben sich Leute überlegt ‚Was brauche ich?‘ und dann wurde das sozusagen nachgebaut. Aber so generisch nachgebaut, dass man trotzdem das Gefühl hatte, ich könnte das anpassen, wenn ich jetzt die Möglichkeiten hätte, so ein bisschen in die Maschine zu gucken, bestimmte Felder umzubenennen oder so.“ (Diskussion M01, 00:10:35)

Auch bei den Python-Skripten im *Jupyter Notebook* wurde die Flexibilität, Funktionen und Parameter genau auf die eigenen Bedürfnisse anzupassen, erkannt. Durch die beliebige Kombination könnten sich bei der Analyse ganz neue Fragestellungen ergeben.

P1: *„Was ich als Positives nochmal heraus streichen möchte ist wirklich diese Variabilität dieser verschiedenen Elemente, die man da kombinieren kann. [...] Wie diese Regler, die man stellen kann und dann schaut, was dabei heraus kommt und dadurch bestimmte Ergebnisse untermauern kann“*

P2: *„Und dann neue Fragestellungen, die sich dann dabei entwickeln und die man natürlich bei einem vorgefertigten Programm dann nicht so einfach ändern könnte.“* (Diskussion M02, 00:12:05)

Allerdings wurde bemängelt, dass für das Ausschöpfen dieser Möglichkeit Python-Kenntnisse erforderlich wären.

„Für diejenigen, die es [Python] schreiben können, sieht das wunderschön wahrscheinlich sogar aus [...]. Wenn man weiß, was da steht, dann ist das glaube ich echt sogar nützlich und man kann überall eingreifen. Aber dann funktioniert's nicht, dann liegt's daran, dass irgendwo eine Leertaste geschrieben ist. Das sind halt so banale Geschichten finde ich.“ (Diskussion M02, 00:11:09)

An dieser Stelle zeigt sich besonders deutlich, wie grundlegend sich die Herangehensweise an die Arbeit mit Texten in der klassischen Geisteswissenschaft gegenüber computergestützter Textanalyse unterscheidet.

Interoperabilität

In Bezug auf *QAnnotate* und *Nodegoat* wurde die fehlende Kompatibilität der Input- und Outputformate der beiden Anwendungen in den Diskussionen besprochen. In diesen Kontext lässt sich auch der geäußerte Wunsch nach einer Komplettlösung einordnen, mit der alle Aufgaben des Arbeitsprozesses gelöst werden können.

„Wonach ich mich sehne, habe ich auch festgestellt, ist wirklich so eine Komplettlösung, wo alle Schritte, die bei der Auswertung anfallen, tatsächlich in einem Programm abwickelbar sind, was dann auch systemisch zusammenhängt. Also dass ich zunächst mal den Datensatz einlese, dann die

Vorarbeiten liefere, dann die Annotationen mache und dann womöglich noch die Visualisierung und das in einem Strang ausführen kann.“ (Diskussion M01, 00:19:48)

Dabei wurde jedoch von Seiten anderer Diskussionsteilnehmenden eingeräumt, dass dieser Ansatz einer einzigen monolithischen Anwendung aus Nutzer*innenperspektive zu anderen Problemen wie Unübersichtlichkeit führen kann. Gleichzeitig wurde auf Probleme aus Perspektive der Entwickler*innen und den alternativen Ansatz von über Standards verknüpfbare Micro-Anwendungen hingewiesen.

„Der Grundpunkt von diesen Micro-Services, oder verschiedenen, kleinen Lösungen ist, dass, wenn man ein Standardformat hat als Output, können andere Forschungsgruppen einen Viewer oder andere Präsentationsformen entwickeln. Dann muss nicht eine Gruppe alle Lösungen finden.“ (Diskussion M01, 00:22:56)

Im Zuge dieser Diskussion ist die Gruppendynamik bei dieser Erhebungsmethodik interessant, da der Teilnehmer, der sich zunächst eine Komplettlösung wünschte, später einräumte, von diesem Ansatz abgekommen zu sein.

4 Reflexion und Ausblick

Die Usability-Analyse mittels Eyetracking und die gewonnenen Erkenntnisse aus den Gruppendiskussionen geben einen ersten Einblick über typische Probleme von Anwendungen im Kontext der Digital Humanities, sowie über Methoden zur Analyse derselben.

(1) Zum einen konnte ermittelt werden, dass elementare Funktionen stets im sichtbaren Bereich der Anwendungen bleiben sollten (siehe *Dillinger* Import). (2) Zum anderen sollten Funktionen zur Analyse bestehender Datensätze idealerweise unmittelbar auffindbar sein und Fehlinterpretationen vermieden werden (siehe *Nodegoat* Filter-Funktion mit durchschnittlich 16,11 Sekunden bzw. 35,13 Sekunden bis zum ersten Klick). (3) Dabei sollte beachtet werden, dass die Ansicht gleichzeitig nicht durch für die Aufgabe weniger zentrale Funktionen unübersichtlich wird (siehe Gruppendiskussion, Kategorie *Verhältnis zwischen Aufwand und Nutzen*).

Die Studie liefert erste Erkenntnisse über die Usability digitaler geisteswissenschaftlicher Werkzeuge, die in folgenden Studien auf zahlreiche Anwendungen übertragen werden können. Die Ergebnisse helfen, Nutzer*innen solcher Anwendungen bei der Entwicklung zu berücksichtigen.

Bibliografie

- [Be16] BENDER, MICHAEL: Forschungsumgebungen in den Digital Humanities, Nutzerbedarf, Wissenstransfer, Textualität. Berlin, Boston : De Gruyter Mouton, 2016 — ISBN 978-3-11-045969-2
- [Bu16] Bulatovic, Natasia; Gnadt, Timo; Romanello, Matteo; Stiller, Juliane; Thoden, Klaus: Usability in Digital Humanities - Evaluating User Interfaces, Infrastructural Components and the Use of Mobile Devices During Research Process. Research

and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science 2016

- [Bu12] Burghardt, Manuel: Usability Recommendations for Annotation Tools. In: 6th Linguistic Annotation Workshop Bd. Proceedings of the ACL. 104-112, 2012
- [Ed12] Edwards, Charlie: The digital humanities and its users. In: (Gold, M. K. Hrsg.): Debates in the digital humanities. University Of Minnesota Press, Minneapolis, 213-232, 2012
- [EW07] Ehmke, Claudia; Wilson, Stephanie: Identifying web usability problems from eye-tracking data. In (British Computer Society Hrsg.): Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1, 119-128, 2007
- [GO12] Gibbs, Fred; Owens, Trevor: Building better digital humanities tools. Digital Humanities Quarterly 6/2012, 2012
- [PB06] Poole, Alex; Ball, Linden J: Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In (Ghaoui, C. Hrsg.): Encyclopedia of Human Computer Interaction, 2006
- [SH10] Schreibman, Susan; Hanlon, Ann M.: Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers. In: Digital Humanities Quarterly 4/2010, 2010
- [Sc03] Schiessl, Michael; Duda, Sabrina; Thölke, Andreas; Fischer, Rico: Eye tracking and its application in usability and media research. MMI-interaktiv Journal, 6/2003, 41-50, 2003
- [St15] Stiller, Juliane; Thoden, Klaus; Leganovic, Oona; Heise, Christian; Höckendorff, Mareike; Gnadt, Timo: Nutzungsverhalten in den Digital Humanities, Aufbau von Forschungsinfrastrukturen für die e-Humanities (Nr. R 1.2.1). DARIAH-DE, 2015
- [Th17] Thoden, Klaus; Stiller, Juliane; Bulatovic, Natasia; Meiners, Hanna-Lena; Boukhe-lifa, Nadia: User-Centered Design Practices in Digital Humanities – Experiences from DARIAH and CENDARI. ABI Technik, 37/2017, 2017
- [Un00] Unsworth, John: Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? In Humanities Computing: formal methods, experimental practice, 2000
- [WX19] Wu, Dan; Xu, Shuang: How Users Gaze and Experience on Digital Humanities Platform?: A Model of Usability Evaluation. In (Taylor, N. G.; Christian-Lamb, C.; Martin, M. H.; Nardi, B. Hrsg.): Information in Contemporary Society. Springer International Publishing, 2019