

Digital Humanities: Laboratorium der Geisteswissenschaften oder Weg nach Atlantis?

Charlotte Schubert (Leipzig)

Spricht man heute von Digital Humanities, so öffnet man damit ein weites Feld, in dem meist eher unklare Vorstellungen davon vorherrschen, was genau damit eigentlich gemeint sei. Sehr gut erkennbar wird dies an den Begriffen, die trotz aller Unterschiedlichkeit deckungsgleich gebraucht werden: So heißt es eben oft nicht ›Digital Humanities‹, sondern ›eHumanities‹, ›Computational Humanities‹ oder auch ›Digitale Geisteswissenschaften‹. Mit welchen unterschiedlichen Konjunkturen diese Begriffe verbunden sind, zeigt eine einfache N-Gramm-Suche mit dem *Google Books Ngram Viewer* (siehe Abb. 1 und 2).¹ Dieses Anschwellen der Begriffe geht einher mit einer intensiv geführten Diskussion darüber, wie die fachliche Zuordnung zu sehen sei: So geht es zum einen um die »Gewinnung und Vermittlung neuen Wissens unter den Bedingungen einer digitalen Arbeits- und Medienwelt«, zum anderen um »Anwendung und Weiterentwicklung von Werkzeugen, [...] Operationalisierung und Beantwortung von Forschungsfragen und [...] Reflexion über die methodischen und theoretischen Grundlagen der Geisteswissenschaften in einer digitalen Welt.«² Verbunden ist mit diesen Definitionsbemühungen ein Zukunftsversprechen: »Grundsätzlich stehen digitalen Geisteswissenschaftlern die gleichen Berufsfelder offen, wie allen anderen Geisteswissenschaftlern auch – allerdings unterscheiden sie sich durch zusätzliche Qualifikationen im methodischen und technischen Bereich und durch ihre zukunftsorientierte Form der geisteswissenschaftlichen Kompetenz vom Gros der Bewerber.«³ Oder: »Die Digital Humanities bilden damit ein eigenständiges Fach mit eigenen, sich ständig weiterentwickelnden Methoden, das einen vollkommen neuen, faszinierenden Blick auf geisteswissenschaftliche Fragestellungen ermöglicht.«⁴

¹ Der *Google Books Ngram Viewer*, <books.google.com/ngrams>, 7.2.2014, erlaubt die Suche nach Wörtern oder Phrasen in den Beständen von *Google Books* und zeigt deren Häufigkeitsverteilung über einen gewählten Zeitraum – z. B. vom Jahr 1800 bis zum Jahr 2008 – an.

² Cologne Center for eHumanities (Hrsg.), *Digitale Geisteswissenschaften*, Köln 2011, <dig-hum.de/sites/dig-hum.de/files/cceh_broschuereweb.pdf>, 7.2.2014, S. 4.

³ Ebd., S. 7.

⁴ Caroline Sporleder, »Was sind eigentlich Digital Humanities?«, in: *Academics.de*, <www.academics.de/wissenschaft/was_sind_eigentlich_digital_humanities_56898.html>, 7.2.2014.

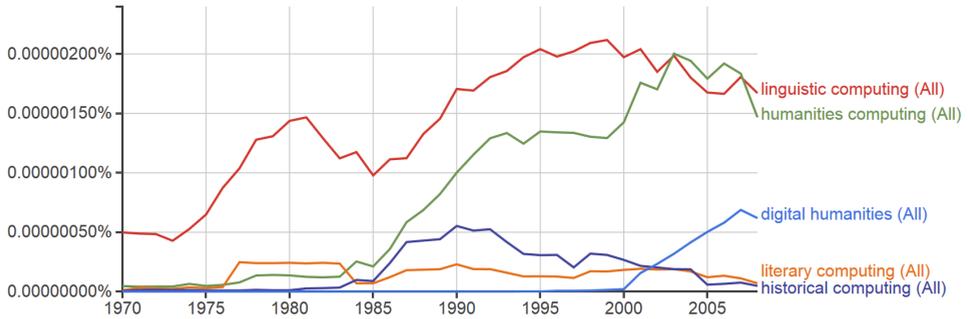


Abbildung 1: Häufigkeit des Auftretens fachspezifischer Bezeichnungen mit ›computing‹ im Vergleich zu ›Digital Humanities‹ im Korpus englischsprachiger Bücher bei *Google Books* von 1970 bis 2008

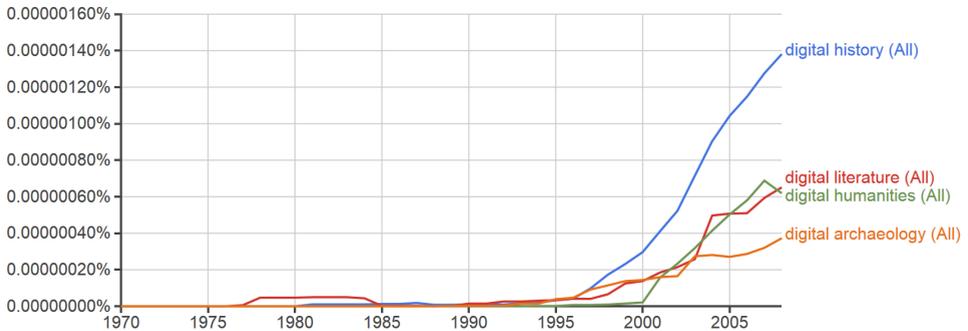


Abbildung 2: Häufigkeit des Auftretens fachspezifischer Bezeichnungen mit ›digital‹ im Vergleich zu ›Digital Humanities‹ im Korpus englischsprachiger Bücher bei *Google Books* von 1970 bis 2008

Nicht alle Forscher und Forscherinnen, die sich in ihrem Arbeitsspektrum mit Projekten befassen, die als ›Digital-Humanities-Projekte‹ bezeichnet werden, teilen diesen Optimismus. So hebt Gerhard Lauer insbesondere die Problematik der von den Digital Humanities geforderten kollaborativen Arbeitsform hervor:

Kollaborative Praxen der Forschung aber passen kaum zu dem genieästhetisch eingefärbten Selbstverständnis des Intellektuellen wie des Gelehrten. Auch das trägt dazu bei, dass Digital Humanities nicht einfach die Zukunft der Geisteswissenschaft sind. [...] Und ob virtuelle Forschungsumgebungen, die Untersuchungsobjekte, Werkzeuge und Publikationen in einer Umwelt zusammenführen, unsere Zukunft auch in den Geisteswissenschaften sind, ist Gegenstand erster Überlegungen. Es spricht vieles dafür, dass auch die Geisteswissenschaften

so selbstverständlich digital werden, wie es andere Wissenschaftsbereiche längst sind. Computer machen auch für die Geisteswissenschaften einen Unterschied, aber welchen, das beginnen wir gerade erst zu entscheiden.⁵

Dieser wohlthuenden Vorsicht stehen aber mittlerweile ganz andere Tendenzen gegenüber, die sich unter den Schlagworten von ›Big History‹ und ›Culturomics‹ fassen lassen: Wie in einer Analogiebildung zu den ›Big Sciences‹, einem Phänomen, das Lorraine Daston als die sich seit dem 19. Jahrhundert entwickelnde Dominanz der Naturwissenschaft als einer Ressourcenbemächtigungspolitik beschrieben hat, kann man die seit einiger Zeit mit ›Big Data‹ und ›Big Data Analytics‹ umschriebenen Phänomene sehen:⁶

We define Big Data as a cultural, technological, and scholarly phenomenon that rests on the interplay of:

- (1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets.
- (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.
- (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.

Insbesondere – befördert und mitfinanziert von *Google* – versucht sich seit einigen Jahren ein Bereich ›Culturomics‹ zu etablieren, der die Methoden von Big Data und Big Data Analytics auf die Geisteswissenschaften überträgt, wie insbesondere der in diesem Kontext entwickelte *Ngram Viewer* zeigt. ›Culturomics‹ – als Parallelbildung zu ›Genomics‹ oder ›Proteomics‹ – soll zum Ausdruck bringen, dass man mit Hilfe solcher Massendatenerhebungen auch in den Geistes- und Kulturwissenschaften zu tragfähigen Analyseergebnissen kommt.

Auf der Ebene der Methodenreflexion stellt sich dies jedoch etwas anders dar: So ist das Zeitalter des Digitalen als ein Zeitalter charakterisiert worden, das vom Erfolg der Kombination aus der Praxis des Narrativen und der Statistik geprägt ist:⁷ »Mit einer modellorientierten Wissenschaftsauffassung, vermutlich sogar mit einer auf Erkennt-

⁵ Gerhard Lauer, »Digital Humanities – die anderen Geisteswissenschaften«, in: *Bulletin der Schweizerischen Akademie der Geistes- und Sozialwissenschaften* 1 (2012), S. 54 f.

⁶ Danah Boyd / Kate Crawford, »Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon«, in: *Information, Communication & Society* 15 (2012), S. 662–679, hier S. 663. Online verfügbar unter < www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878 >, 3.12.2014.

⁷ Ben Kaden, »In weiter Ferne und nah. Über Dekonstruktion, Digitalkultur und Digital Humanities«, in: *LIBREAS. Library Ideas*, < libreas.wordpress.com/2013/06/27/dekonstruktion_humanities >, 7.2.2014.

nis (statt Verständnis) und Kontrolle (statt Anerkennung) gerichteten Wissenschaft lässt sich die Dekonstruktion kaum in Einklang bringen.« Kaden betont hierbei die Bedeutung einer »Praxis der Nähe« (im Hinblick auf Textarbeit) und die Gefahren, die durch die Massendigitalisierung zu Untergang und Ende des Kanons führen könnten.

Dies ist eine bedenkenswerte Überlegung, die die von Franco Moretti mit dem Gegensatz von ›Distant Reading‹ und ›Close Reading‹ so eindrücklich beschriebene Methodenreflexion aufgreift. Moretti hat bekanntlich gezeigt, welches Potential quantitative Analysemodelle in Verbindung mit graphischen Visualisierungen haben können.⁸

Unter dem Schlagwort ›Distant Reading‹ sind seit einigen Jahren neue Möglichkeiten diskutiert worden, aus der schiereren Menge von Daten durch die Anwendung algorithmenbasierter Auswertung mit den Methoden etwa des Text Mining, des Clustering oder des Topic Modeling neue Zusammenhänge aus sehr großen Text- und Datenmengen zu erkennen.⁹ Komplexitätsreduktion, Visualisierung und exploratives Experimentieren haben aber auch auf ganz neue Fragen geführt, von denen nicht zuletzt die Qualität der zugrunde gelegten Daten eine wesentliche ist. Diese prägnanten Ausführungen zur Taxonomie der Formen in Morettis *Graphs, Maps, Trees* lassen sich verallgemeinern und auch auf ganz andere Visualisierungsformen übertragen, die nicht nur für das neue Feld der Digital Humanities aufgrund ihrer Verbindung ganz unterschiedlicher disziplinärer Traditionen von Bedeutung sind. Die Fragen, die dabei im Vordergrund stehen, sind: Welche Art von Information wird verwendet, wie wird sie verarbeitet, welche Formalisierungen werden eingesetzt? Vor allem: Welche impliziten Bedeutungen werden mitgetragen? Sowie schließlich: Wie verhält sich dieser Ansatz zu dem ›Close Reading‹, wie es sich beispielsweise in der traditionellen, historisch-philologischen Textanalyse etabliert hat?

Als erster Ansatzpunkt soll hier der Begriff der ›Daten‹ thematisiert werden. Grundlage aller Arbeiten im Bereich der Digital Humanities ist die Strukturierung, Beschreibung und Organisation von Daten. Dabei werden zahlreiche Begriffe verwendet, um

⁸ Franco Moretti, *Graphs, Maps, Trees. Abstract Models for Literary History*, London-New York 2007. Insbesondere Gary S. Schaal (Hamburg) und Roxana Kath (Leipzig) haben hierzu wegweisende Überlegungen zur Etablierung einer Neuen Visuellen Hermeneutik entwickelt: Gary S. Schaal u. a., »New Visual Hermeneutics«, in: *Automatisierte Textanalyse. Sonderheft der Zeitschrift für Germanistische Linguistik*, 2015 (i. V.). – Roxana Kath, »Aesthetics are (ir)relevant: Überlegungen zu einer neuen visuellen Hermeneutik in den Geisteswissenschaften«, in: *Zeitschrift für Politische Theorie* 5 (2014), S. 97–120.

⁹ Zur Methodendiskussion vgl. v. a. Lev Manovich, *The Language of New Media*, Cambridge 2001, und Gregory Crane, »What Do You Do with a Million Books?«, in: *D-Lib Magazine* 12 (2006), < www.dlib.org/dlib/march06/crane/03crane.html >, 26.4.2013.

Daten in ihrer Begrifflichkeit zu umschreiben, etwa ›Information‹, ›Zeichen‹ oder ›String‹. Für die Strukturierung werden ›Schemata‹ definiert und für die Organisation, das heißt die Archivierung, Erschließung und Bereitstellung, weitere Techniken, Apparaturen und ganze Zyklen von Prozeduren. Für den vielleicht wichtigsten Begriff in diesem Zusammenhang, nämlich ›Daten‹, gibt es nun allerdings keine einheitliche, allgemein akzeptierte Definition: Man kann darunter ›Rohdaten‹ verstehen, aber auch verarbeitete Daten oder ganz positivistisch: Zahlen.¹⁰ Auch die ebenso gern verwendete Gleichsetzung von Daten mit Bits oder Bytes zeigt das begriffliche Problem schon an: Bits und Bytes sind das Ergebnis eines schaffenden Vorgangs, sie sind kein Ergebnis einer Beobachtung.¹¹ Insofern können ›Daten‹ als technische Artefakte bezeichnet werden, die in einem jeweils spezifischen, kulturellen, historischen und sozialen Zusammenhang stehen.

Der gleiche Befund ist für die sogenannten ›Metadaten‹ festzustellen: Metadaten, so eine prägnante Formulierung von Floridi »are indications about the nature of some other (usually primary) data. They describe such as location, format, updating, availability, usage restrictions, and so forth.«¹² Die Unbestimmtheit in »and so forth« ist allerdings bezeichnend. Am Beispiel der Gegenüberstellung einer algorithmenbasierten, quantitativen Auswertung und einer – ebenfalls algorithmenbasierten – semantischen Auswertung antiker Texte soll hier nun demonstriert werden, wie sich diese Probleme in der praktischen Arbeit und Anwendung niederschlagen, aber auch welche Lösungsstrategien für diese methodischen Fragen möglich sind.

Den klassischen Altertumswissenschaften kommt in dem Konzert der digitalen Geisteswissenschaften meines Erachtens eine besondere Rolle zu, weil sie früher als andere und daher heute in komfortabelster Weise über fast vollständig digitalisierte Textkorpora verfügen. Sie können aufgrund dieser besonderen Situation als *Fall par excellence* dafür präsentiert werden, wie man die methodischen Probleme benennen und eingrenzen kann.

Wie ich dies praktisch meine, möchte ich exemplarisch erläutern am Beispiel der Arbeiten, die im Rahmen des Projekts *eAQUA* durchgeführt wurden,¹³ und zwar an einer experimentell entwickelten Analyse zu dem Begriff ›Atlantis‹.

¹⁰ Vgl. dazu ausführlich Jacob Voß, *Describing Data Patterns. A general deconstruction of metadata standards*, Diss. HU Berlin 2013, <urn:nbn:de:kobv:11-100212118>, 9.2.2014, S. 43.

¹¹ Ebd. S. 42. Voß diskutiert in diesem Zusammenhang auch die zeichentheoretischen Grundlagen nach Peirce und de Saussure in den heutigen Datenmodellen.

¹² Luciano Floridi, *Information – a very short introduction*, Oxford 2010, S. 23.

¹³ *Extraktion von strukturiertem Wissen aus Antiken Quellen für die Altertumswissenschaft*, <www.eaqua.net>, 3.12.2014.

1. Close Reading

Für einfache oder kombinierte Wortsuchen stehen heute in den Altertumswissenschaften bereits einige etablierte Programme zur Verfügung.¹⁴ Die Suchfunktionen neuerer Text-Mining-Verfahren gehen jedoch darüber hinaus, indem sie die Möglichkeit bieten, einen syntagmatischen Kontext anzuzeigen. Ein syntagmatischer Kontext liegt vor, wenn zwei Wortformen insofern in einer Relation stehen, als dass sie gemeinsam auftreten, das heißt in mindestens einem lokalen Kontext, der beide Wortformen enthält. Darauf beruht die Suchfunktion von eAQUA und Ergebnisse dieser Art werden im eAQUA-Portal bei der Eingabe eines Wortes oder einer Wortform in die Suchmaske sowohl in Form eines Graphen visualisiert (graphische Darstellung eines Wortnetzes von Satzkookkurrenten) als auch mit signifikanten Kookkurrenzen und Nachbarschaftskookkurrenzen (jeweils nach der Stellung sortiert) angezeigt. Mit Hilfe einer statistischen Kookkurrenzanalyse, berechnet mit der Log-Likelihood-Ratio, wird die Häufigkeit des gemeinsamen Auftretens eines Wortpaares im Verhältnis zum Gesamtvorkommen im Textkorpus angegeben, so dass systematische Kookkurrenzprofile erstellt und visualisiert werden (siehe Abb. 3 und 4).

Nach dem Mythos, wie Hesiods *Theogonie* ihn überliefert, kann der Name Atlantis auf Atlas als einen der Titanen, Sohn von Uranos und Gaia, zurückgeführt werden (siehe Abb. 4, Beleg Nr. 4, Hesiod, Theog. 938 zur Ἀτλαντὶς Μαίη). Atlas soll auch einen Aufstand gegen Zeus (Ζῆν) angeführt haben und mit dem Tragen des Himmelsgewölbes bestraft worden sein. Platon leitet später von Atlas den Namen Ἀτλαντὶς ab (Kritias 114a). Etymologisch ist in Ἀτλαντὶς wohl τλῆναί = »(er)tragen« zu sehen.

Die eigentliche Überlieferung zu Atlantis, wie sie heute mit dem Namen verbunden wird, geht auf Platon und seine Werke *Timaios* und *Kritias* zurück (Timaios 24e–25d und Kritias 113b–121c). Platon lässt Kritias von einem Dichtwerk berichten, das sein Großvater noch vom alten Solon gehört habe: Dieser habe aus Ägypten die Geschichte von Atlantis mitgebracht, aber aufgrund diverser politischer Unruhen nicht fertigstellen können. Den Inhalt – die Konstruktion der Stadt, die Organisation der Gesellschaft sowie die Geschichte des Untergangs in einer gewaltigen Flut (κατακλυσμός, siehe Abb. 5) – berichtet Kritias aber auch in dem gleichnamigen Dialog Platons und diese Fassung ist seither in ihren Grundzügen kanonisch geworden und daher mehr als bekannt.

¹⁴ Zum Beispiel das Programm *Diogenes* von Peter J. Heslin, die Online-Suchfunktion im *Thesaurus Linguae Graecae* (TLG), < www.tlg.uci.edu >, 3.12.2014, oder *Pandora*. Das Programm *Pandora* ist eine Retrieval-Software für Hypercard, die zur Recherche in der CD-ROM-Version des TLG (Version D und E) verwendet werden konnte. Es kann noch heute heruntergeladen werden unter < e3.uci.edu/00w/18130/Pandora2.5.3.hqx >, 14.12.2014.

Belegstellen für Ἀτλαντῆς und Ζηνὶ

Anzahl 5 (Ἀτλαντῆς : 16 , Ζηνὶ : 212)

Gefundene Belegstellen

1. <Ζηνὶ ὄ' ἄρ'> Ἀτλαντῆς Μαίη τέκε κούδιμον Ἐρμῆν, κήρυκ' ἄθανάτων, ἱερὸν λέχος εἰσαναβάσσα.
Source: ARSENIUS *Paroemiogr. (9018) (Monembasiensis - A.D. 15-16)*: *Apophthegmata*
Video: NULL
Publication: *Apophthegmata*, ed. E.L. von Leutsch, *Corpus paroemiographorum Graecorum*, vol. 2. Göttingen: Vandenhoeck & Ruprecht, 1851 (repr. Hildesheim: Olms, 1958): 240-744. (Cod: 23.388: *Paroem.*)
Document citation: //8/35k/1 to //8/35k/2 (Schema: *Centuria/section/line*)
2. Ζηνὶ ὄ' ἄρα Ἀτλαντῆς Μαίη τέκε κούδιμον Ἐρμῆν, κήρυκ' ἄθανάτων, ἱερὸν λέχος εἰσαναβάσσα:
Source: CLEMENS ALEXANDRINUS *Theol. (0555) (Alexandrinus - A.D. 2-3)*: *Stromata*
Video: Scholia: Cf. SCHOLIA IN CLEMENTEM ALEXANDRINUM (5048)
Publication: *Stromata*, ed. O. Stählin, L. Früchtel and U. Treu, *Clemens Alexandrinus*, vols. 2, 3rd edn. and 3, 2nd edn. [Die griechischen christlichen Schriftsteller 52(15), 17. Berlin: Akademie-Verlag, 2:1960; 3:1970]: 2:3-518; 3:3-102. *Lib. 1-6: vol. 2. *Lib. 7-8: vol. 3. (Cod: 166,077: *Phil., Theol.*)
Document citation: /1/211/105/5/311 to /1/211/105/5/411 (Schema: *Book/chapter/section/subsection/line*)
3. Ζηνὶ ὄ' ἄρ' Ἀτλαντῆς Μαίη τέκε κούδιμον Ἐρμῆν, κήρυκ' ἄθανάτων, ἱερὸν λέχος εἰσαναβάσσα.
Source: EUSEBIUS *Scr. Eccl. Theol. (2018) (Caesariensis - A.D. 4)*: *Praeparatio evangelica*
Video: NULL
Publication: *Praeparatio evangelica*, ed. K. Mras, *Eusebius Werke*, Band 8: *Die Praeparatio evangelica* [Die griechischen christlichen Schriftsteller 43.1 & 43.2. Berlin: Akademie-Verlag, 43.1:1954; 43.2:1956]: 43.1:3-613; 43.2:3-426. (Cod: 241,486: *Apol.*)
Document citation: /10/12/20/3t to /10/12/20/4t (Schema: *Book/chapter/section/line*)
4. Ζηνὶ ὄ' ἄρ' Ἀτλαντῆς Μαίη τέκε κούδιμον Ἐρμῆν, κήρυκ' ἄθανάτων, ἱερὸν λέχος εἰσαναβάσσα.
Source: HESIODUS *Epic. (0020) (Ascræus - 8/7 B.C.?)*: *Theogonia*
Video: Cf. et AEGIMIUS (0668) *Vita: Cf. VITAE HESIODI PARTICULA (1749) Scholia: Cf. SCHOLIA IN HESIODUM (5025)*
Publication: *Theogonia*, ed. M.L. West, *Hesiod: Theogony*. Oxford: Clarendon Press, 1966: 111-149. (Cod: 6,969: *Epic.*)
Document citation: ///938a to ///939a (Schema: *Line*)
5. <Ζηνὶ ὄ' ἄρ' Ἀτλαντῆς Μαίη>.
Source: Joannes GALENIUS *Gramm. (3039) (Constantinopolitanus - A.D. 12)*: *Allegoriae in Hesiodi theogoniam*
Video: NULL
Publication: *Allegoriae in Hesiodi theogoniam*, ed. H. Flach, *Glossen und Scholien zur hesiodischen Theogonie*. Leipzig: Teubner, 1876 (repr. Osnabrück: Biblio Verlag, 1970): 295-365. (Cod: 20.301: *Schol.*)
Document citation: //358/24 to //358/24 (Schema: *Page/line*)

Abbildung 4: Belegstellen für das kookkurrente Auftreten von Ἀτλαντῆς und Ζηνὶ über die Funktion ›Double Quotation‹ aus dem Ergebnis für Ἀτλαντῆς mit der eAQUA-Kookkurrenzsuche und der Visualisierung in Abb. 3

Die Voraussetzungen der Kookkurrenzanalyse sind bestimmte Schritte des Text Preprocessing wie beispielsweise Satzsegmentierung und Worttokenisierung. Verwendete Ordnungsprinzipien wie ›Satz‹, ›Nachbarschaft‹ etc. werden dabei jedoch ohne Berücksichtigung der Spezifika der Textgenese in ihrer ca. 2000-jährigen Editions-geschichte angewandt, ebensowenig wie die dabei entstandene Graphienvarianz mit entsprechenden Änderungen der Schreibkonventionen Eingang findet. Weiterhin wird in der Regel in digitalen Volltextkorpora antiker Texte aus nachvollziehbarem Grund nur je eine – mehr oder weniger moderne – Edition eines antiken Textes aufgenommen, da mehrere Editionen ein und desselben Textes zu verfälschten Analyseergebnissen führen würden.

Praktisch heißt dies aber, dass die in der Regel modernen, textkritischen Eingriffe in die Texte, die meist ebenso modernen Schreibkonventionen wie zum Beispiel die Setzung eines Punktes, die Vereinheitlichungen (Iota adscriptum zum Iota subscriptum) oder auch die Auflösung der Scripta continua in ›Strings‹ zu einer Textgestalt

Belegstellen für Ἀτλαντὶς und κατακλυσμών

Anzahl 5 (Ἀτλαντὶς : 16 , κατακλυσμών : 25)

Gefundene Belegstellen

1. ἡ δὲ Ἀτλαντὶς νῆσος, ἅμα Λιβύης καὶ Ἀσίας μεζώνων, ἧ φησὶν ἐν Τιμαίῳ Πλάτων, ἡμέρα μὴ καὶ νυκτὶ σεισμῶν ἐξαισίων καὶ κατακλυσμών γενομένων δόσα κατὰ τῆς θαλάττης ἐξαίφνης ἠφάνισθη, γενομένη πέλαγος, οὐ πλωτῶν, ἀλλὰ βαρᾶθρῶδες.
Source: PHILO JUDAIEUS Phil. (0018) (Alexandrinus - 1 B.C.-A.D. 1): De aeternitate mundi
Vide: NULL
Publication: De aeternitate mundi, ed. L. Cohn and S. Reiter, Philonis Alexandrini opera quae supersunt, vol. 6. Berlin: Reimer, 1915 (repr. De Gruyter, 1962): 72-119. (Cod: 9,401: Phil.)
Document citation: //4/1411/111 to //4/1421/111 (Schema: Section/line)
2. ὕστερῳ δὲ χρόνῳ σεισμῶν ἐξαισίων καὶ κατακλυσμών γενομένων, μίσης ἡμέρας καὶ νυκτὸς χαλεπῆς ἐπέλθοῦσης, τὸ τε παρ' ὕμιν μάχιμον πᾶν ἄθροον ἔδω κατὰ γῆς, ἡ τε Ἀτλαντὶς νῆσος ὡσαύτως κατὰ τῆς θαλάττης δόσα ἠφάνισθη:
Source: PLATO Phil. (0059) (Atheniensis - 5-4 B.C.): Timaeus
Vide: Cf. et SOCRATICORUM EPISTULAE (0637) Lexicon: Cf. TIMAEUS Sophista Gramm. (2602) Scholia: Cf. SCHOLIA IN PLATONEM (5035)
Publication: Timaeus, ed. J. Burnet, Platonis opera, vol. 4. Oxford: Clarendon Press, 1902 (repr. 1968): St III.17a-92c. (Cod: 24,104: Dialog., Phil.)
Document citation: //25/c/6 to //25/d/3 (Schema: Stephanus page/section/line)
3. ἡ δὲ Ἀτλαντὶς νῆσος, ἅμα Λιβύης καὶ Ἀσίας μεζώνων, ἧ φησὶν ἐν Τιμαίῳ Πλάτων, ἡμέρα μὴ καὶ νυκτὶ σεισμῶν ἐξαισίων καὶ κατακλυσμών γενομένων δόσα κατὰ τῆς θαλάττης ἐξαίφνης ἠφάνισθη, γενομένη πέλαγος, οὐ πλωτῶν, ἀλλὰ βαρᾶθρῶδες.
Source: POSIDONIUS Phil. (1052) (Apamensis Rhodius - 2-1 B.C.): Fragmenta
Vide: NULL
Publication: Fragmenta, ed. W. Theiler, Posidonios. Die Fragmente, vol. 1. Berlin: De Gruyter, 1982: 16-72, 75-238, 242-244, 255-269, 279-280, 285, 290-297, 307, 310-346, 348-350, 354-357, 375-386. Dup. partim 1052 003. (Q: 106,815: Geogr., Hist., Math., Nat. Hist., Phil., Tact.)
Document citation: //310a/145n to //310a/148n (Schema: Fragment/line)
4. <Υστερῳ δὲ χρόνῳ σεισμῶν ἐξαισίων καὶ κατακλυσμών γενομένων μίσης ἡμέρας καὶ νυκτὸς χαλεπῆς ἐπέλθοῦσης τὸ τε παρ' ὕμιν μάχιμον πᾶν ἄθροον ἔδω κατὰ γῆς, ἡ τε Ἀτλαντὶς νῆσος ὡσαύτως κατὰ τῆς θαλάττης δόσα ἠφάνισθη.
Source: PROCLUS Phil. (4036) (Atheniensis - A.D. 5): In Platonis Timaeum commentaria
Vide: NULL

Abbildung 5: Belegstellen für das kookkurrente Auftreten von Ἀτλαντὶς und κατακλυσμών über die Funktion ›Double Quotation‹ aus dem Ergebnis für Ἀτλαντὶς mit der eAQUA-Kookkurrenzsuche und der Visualisierung in Abb. 3

führen, die schon ihrerseits eine moderne Konstruktion darstellt. Auch daraus ergibt sich das große Problem der falsch erkannten (›Precision‹) und der nicht gefundenen (›Recall‹) Belege und Zitate/Sätze, für die kaum eine verlässliche Quote angegeben werden kann. Der hier auch auf mehreren Ebenen offenbar werdende hermeneutische Zirkel sollte zumindest darauf hinweisen, dass, wenn wie zum Beispiel bei ›String‹ die Definitionsgrundlage modern ist und die Textgrundlage auf modernen Editionsprinzipien beruht, die Ergebnisse einer Analyse mit Hilfe des Text Mining auf einer Grundlage basieren, die modern und vor allem nicht reversibel ist – im Unterschied zu einer Textarbeit mit herkömmlichen, textkritischen Printeditionen, deren Apparate als Apparatus criticus und Apparatus fontium eben diese editorischen Eingriffe sowie die textkritischen Differenzen zu älteren Editionen kenntlich machen.

2. Distant Reading

Im Folgenden soll beschrieben werden, wie sich das angedeutete Problem in den Bereich des ›Distant Reading‹ einordnen lässt, indem die quantitative Auswertung eines großen Textkorpus anhand der Metadaten, hier des TLG-E, demonstriert wird. Für

die Visualisierung einer quantitativen Auswertung ist auf der Grundlage der Metadaten zu Autorennamen, Orten ihrer Wirkung sowie der chronologischen Einordnung mit Hilfe des Programms *Gephi* eine solche Auswertung durchgeführt worden (siehe Abb. 6). *Gephi* ist ein graphbasiertes Datenerhebungs- und Visualisierungswerkzeug und wertet statistische Informationen, inklusive der Beziehungen, aus und ermöglicht eine graphische Repräsentation, die typische Faktoren für die Netzwerkanalyse, wie beispielsweise ›Betweenness‹, Dichte und den Clusterkoeffizienten ermittelt. Die ›Groß-Visualisierung‹ mit Hilfe von *Gephi* zeigt sämtliche Autorennamen, deren zugehörige Orte und Epochen jeweils als Knoten in einer einzigen Graphik an.

Die auffällige Konzentration auf den Knoten ›Atlantis‹ verweist auf eine Besonderheit der antiken Texte: Viele antike Texte sind namentlich nicht zuzuweisen beziehungsweise werden als Fragmente von modernen Editoren antiken Autoren zugewiesen. Diese komplizierte Textsituation, die der statistischen Auswertung der Autorenschaft entgegensteht, wird zudem von einer weiteren Problematik überschattet: Wir kennen von sehr vielen antiken Autoren weder die Herkunfts-, geschweige denn die Wirkungsorte. Insofern wäre, unter korrekter Berücksichtigung dieses Sachverhaltes, eine Netzwerk-Visualisierung wie die hier vorliegende, eigentlich unmöglich. Um jedoch trotz allem zu einer Anwendung zu kommen, ist in der vorliegenden Datentabelle für die unbekanntenen Ortszuweisungen der Name ›Atlantis‹ und für diesen ein fiktiver Ort im Mittelmeer gewählt worden.¹⁵

Auf den ersten Blick scheint dieses Vorgehen auf einer willkürlichen Konstruktion zu beruhen, in der mit der Kategorie ›Atlantis‹ eine Benennung und Lokalisierung gewählt wurde, die sich einreihet in die unendliche Kette der Geschichten über Atlantis, die ausgehend von dem platonischen Mythos seit mehr als 2 000 Jahren zu vielen phantasievollen Texten geführt hat.¹⁶

Bei näherer Betrachtung zeigt sich, dass es sich bei den Atlantis-Geschichten um einen Wissensdiskurs des Imaginären handelt, der, wie andere Wissensdiskurse auch, Ordnungsprinzipien aufweist.¹⁷ Bekanntlich klassifizieren, sortieren, kategorisieren und segmentieren Wissensdiskurse auf ihre je spezifische Weise, und hier schließt sich die Überlegung an, inwiefern solche Klassifizierungen in unserer Text-Mining-Suche des ›Distant Reading‹ auf vorgeschaltete Ordnungsprinzipien zurückgehen.

¹⁵ Die Idee dazu hatte meine Mitarbeiterin Dr. Roxana Kath, der an dieser Stelle noch einmal herzlich für diese wagemutige Konstruktion gedankt sei.

¹⁶ Zu Atlantis siehe Umberto Eco, *Die Geschichte der legendären Länder und Städte*, München 2013, S. 183–221.

¹⁷ Michel Foucault, *Die Ordnung der Dinge. Eine Archäologie der Humanwissenschaften*, Frankfurt 1971, S. 165 ff.

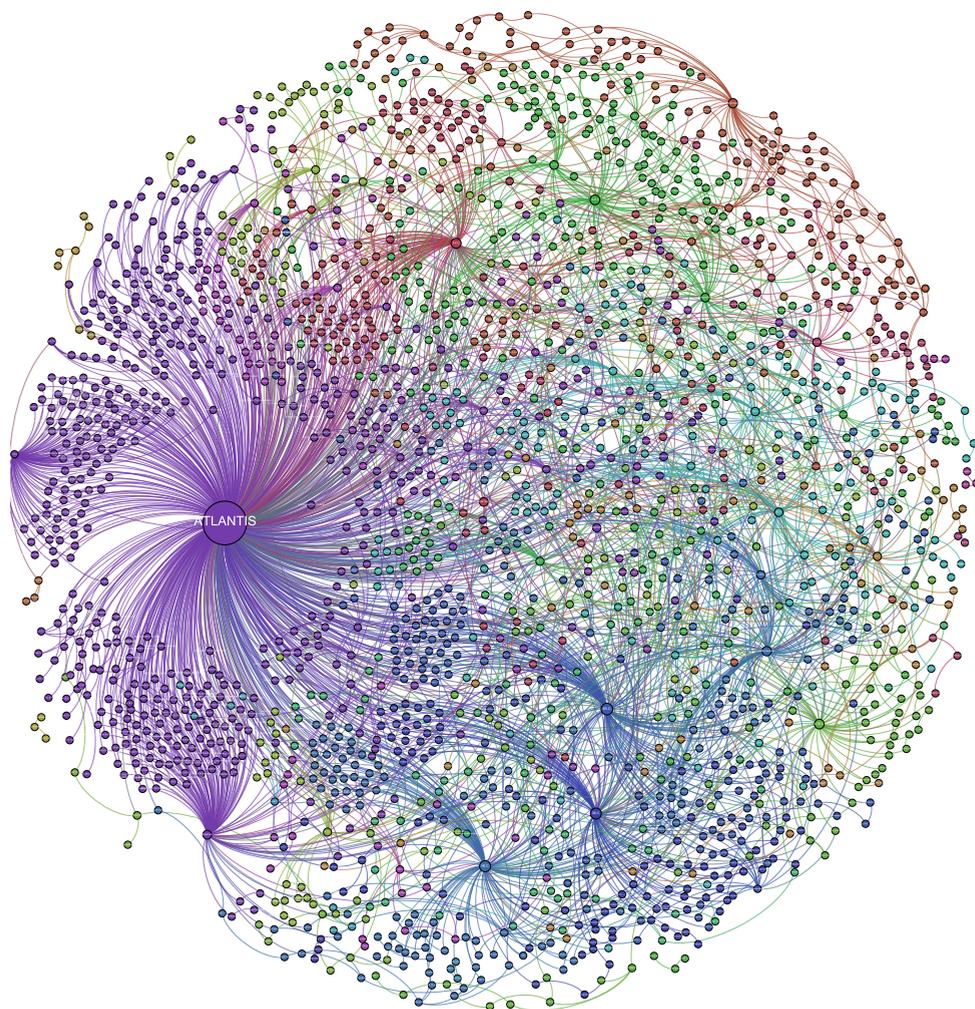


Abbildung 6: Mit *Gephi* erstellte Visualisierung auf der Basis des Metadatenatzes (Autoren-
namen, Orte, Epochen) des TLG-E. Das farbige, vergrößerbare Original ist auf www.eaqua.net unter ›Dokumentation/Anwendungsbeispiele‹
zu finden.

3. Fazit

Metadaten-Schemata können heute als Ordnungsprinzipien eines Wissensdiskurses aufgefasst werden, die den von Foucault herausgearbeiteten Klassifizierungs- und Kategorisierungsschemata aus der Zeit der beginnenden Naturwissenschaften vergleichbar sind: Je nach verwendetem Schema, sei es CIMDI aus CLARIN-De, CIDOC-CRM oder Metadaten nach Dublin Core, werden andere Kategorien verwendet, zu deren Deckungsfähigkeit eigene Interoperabilitätsprogramme entwickelt werden müssen.¹⁸ Die Ontologien, auf denen die Metadaten schemata basieren, sind unterschiedlicher Art – bedingt durch die Gattungen, Objekte und Genres, auf deren Klassifizierung sie angewendet werden sollen und für deren Erschließung sie entwickelt wurden. So verweist CIMDI auf seinen linguistischen Hintergrund, CIDOC-CRM auf den objektbezogenen und musealen Kontext und der Dublin-Core-Standard auf die Erfordernisse des Bibliotheksbereichs. Daraus ergibt sich eine gewisse Kontingenz, die durchaus mit lexikographischen Ordnungsverfahren vergleichbar ist: Deren Etikettierungen sind ebenfalls begriffsbezogene Datenkategorien, die sich aus fachspezifischen und in dieser Hinsicht autonom geprägten Voraussetzungen speisen. Auch das abstrakteste Meta-Modell zur Systematisierung solcher Benennungen wird die Kontingenz dieser Voraussetzungen nicht vermeiden oder beseitigen können. So stellt sich im Ergebnis eine deutliche Parallelität im Vorgehen dar: Die scheinbar hochgradig konstruierte Verwendung des imaginären ›Atlantis‹ als Label oder Kategorie für nicht bekannte Orte ist keineswegs eine Fiktionalisierung, vielmehr handelt es sich um die Einführung eines Ordnungsprinzips, das letztlich nicht weniger konstruiert ist als andere moderne Zuweisungen, wie wir sie auch in diversen Metadatenmodellen erkennen. Insofern sind, um den Blick wieder zurück auf die Ausgangsfrage der algorithmenbasierten Analyse eines Textkorpus zu lenken, zwei Ebenen zu berücksichtigen: diejenige der Kontingenz der Kategorien der Metadaten und diejenige der Kontingenz der Auszeichnungskategorien. Beide beruhen auf hochgradig konstruierten Diskursen und insofern befinden wir uns heute gewiss in einer dem Laboratorium ähnlichen Situation, indem wir – durchaus hypothesengeleitet – experimentieren. Gleichzeitig bietet das neue Feld der Digital Humanities, wie unsere Atlantis-Konstruktion zeigt, auch die Chance, in einer spielerischen Form des Konstruktivismus neue Wege zu gehen, die nur den Geisteswissenschaften möglich sind.

¹⁸ Abruf 21.8.2014: »Component Metadata«, in: *Clarín*, <www.clarin.eu/content/component-metadata>; *The CIDOC Conceptual Reference Model*, <www.cidoc-crm.org>; *Dublin Core Metadata Initiative*, <dublincore.org>. Vgl. Irene Lourdi u. a., »Semantic Integration of Collection Description. Combining CIDOC/CRM and Dublin Core Collections Application Profile«, in: *D-Lib Magazine* 15 (2009), <www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>.