

Rechtsgeschichte Legal History

www.rg.mpg.de

<http://www.rg-rechtsgeschichte.de/rg24>

Rg **24** 2016 344 – 346

Jörg Wettlaufer

Digitale Transformation der Geschichtswissenschaft: Welcher Erkenntnisfortschritt? Wann?

Dieser Beitrag steht unter einer
Creative Commons cc-by-nc-nd 3.0



Jörg Wettlaufer

Digitale Transformation der Geschichtswissenschaft: Welcher Erkenntnisfortschritt? Wann?

Welche Forschungsergebnisse lassen Digital Humanities erwarten? Um seriös über solche Dinge sprechen zu können, müsste man zunächst einmal definieren, was überhaupt ein Erkenntnisfortschritt in der Geschichtswissenschaft wäre. Über das, was hier als Fortschritt zu betrachten sei, ist sich die »Zunft« bislang nämlich keineswegs einig.¹ Ohne diese Diskussion in diesem Rahmen aufnehmen zu können, möchte ich für das hier vorgetragene begrenzte Anliegen unter Erkenntnisfortschritt eine Erkenntnis oder eine Auffassung hinsichtlich einer historischen Detailfrage verstehen, die neu in die Fachliteratur Eingang findet und maßgeblich auf dem Einsatz von Methoden der Informatik und Statistik an historischen Quellen beruht, wie sie heute in den sog. Digital Humanities Anwendung finden. Es geht hier um digitale Methoden, die sozusagen brandneu seit etwa zehn Jahren die Angebotspalette der Werkzeuge des Historikers (frei nach der klassischen Proseminarlektüre des Ahasver von Brandt) erweitern und sowohl dabei helfen können Texte zu analysieren als auch sie in einer Weise aufzubereiten, die neue Erkenntnisse, also einen echten Fortschritt in der Forschung, erwarten lassen.² Je nach Epoche unterscheiden sich Werkzeuge und Erkenntnismöglichkeiten natürlich erheblich – ich beziehe mich vor allem auf das mir vertraute Mittelalter und die Frühe Neuzeit. Antike und Zeitgeschichte haben, selbstredend, andere Schwerpunkt und entsprechend angepasste Werkzeuge.

Ob die oben gestellte Frage nach dem Erkenntnisfortschritt durch die digitale Transformation der Geschichtswissenschaft nun positiv oder negativ zu beantworten ist, hängt ganz von der Forschungsfrage ab, die man an die Quellen stellen möchte. Für den überwiegenden Teil der klassischen Fragen

auf abstrakter Ebene (z. B. nach den Voraussetzungen der föderalen Struktur Deutschlands) erbringen digitale Methoden zunächst einmal rein gar nichts außer einer allerdings erheblichen Arbeits erleichterung und Beschleunigung hinsichtlich der Suche und der Beschaffung von Einzelinformationen. Ich will dies kurz an einem (eigenen) Beispiel konkret erläutern – nämlich Forschungen zur mittelalterlichen und frühneuzeitlichen Strafrechtsgeschichte. Dieses Projekt fragt, ausgehend von einer anthropologischen Fragestellung hinsichtlich des sozialen Gebrauchs von Scham in historischen Gesellschaften, nach Transformationsprozessen, die sich im Gebrauch von beschämenden Strafen seit dem hohen Mittelalter in Europa beobachten lassen. Zur Beantwortung dieser Frage sind eine Vielzahl von Quellen in verschiedenen europäischen Sprachen notwendig, denn das Phänomen der Verwendung dieser Strafen im entstehenden öffentlichen Strafrecht ist nicht regional, sondern zumindest ein (west-)europäisches, und es bieten sich sogar Vergleiche mit anderen, außereuropäischen Kulturkreisen an, um die europäische Praxis besser einordnen zu können.³ Bei einer solchen Forschungsfrage gibt es kein geschlossenes Korpus von Texten, auf das man zurückgreifen könnte (und wenn es dies gäbe, ließe sich die Frage vielleicht schnell beantworten). Vielmehr entsteht ein solches Korpus erst im Laufe der Forschungsarbeit und wird dabei durch die zumeist händische Kodierung und Anreicherung mit Metadaten erst auswertbar gemacht. Hierbei können z. B. Datenbanken und digitale Tools zur Speicherung und schnellen Bereitstellung von Informationen hilfreich sein, aber der Erkenntnisfortschritt wird erst in der Zusammenschau und der Kontextualisierung des Materials erreicht – einem Prozess, der momentan noch nicht digital

1 PARAVICINI (2010).

2 WETTLAUFER/WESTPHAL (2014).

3 SÈRE/WETTLAUFER (2013). Siehe auch unter www.shamestudies.de (1.4.2016).

nachzubilden ist, da er den vollen Kontext menschlicher Erfahrungsräume erfordert – eine Komplexität also, die noch keine Ontologie oder neuronales Netzwerk außerhalb des menschlichen Gehirns heute in der Lage ist zu liefern.

Es sind somit zurzeit vor allem die Fragestellung und die zur Beantwortung der Frage notwendigen Quellen, die über die Möglichkeit des effektiven Einsatzes von digitalen Werkzeugen entscheiden. Der Rest ist dem weiten Feld der Datenpräsentation zuzurechnen, der allerdings eine wichtige Aufgabe in der Verbreitung und Kommunikation von Forschungsergebnissen zukommt. Hinsichtlich der Quellen ist besonders eine Voraussetzung für den Einsatz von Werkzeugen der Computerlinguistik und statistischen Textanalyse wie z. B. Topic Modeling zu beachten: Die Texte oder Bilder müssen als Korpus beschreibbar sein, möglichst digital erkennbare Muster aufweisen und vor allem in einer Menge vorliegen, die den Einsatz von automatisierter Datenverarbeitung sinnvoll macht. Die meisten Algorithmen, abgesehen vielleicht von einigen statistischen Funktionen, entfalten ihr eigentliches Potential eben erst im Bereich von Text- oder Bildmengen, die eine Bearbeitung durch einen Menschen nicht mehr zulassen (distant reading).

Man sollte bzw. müsste also die Frage nach dem digitalen Erkenntnisfortschritt für geschichtswissenschaftliche Forschung noch etwas präzisieren. Welchen Erkenntnisfortschritt lassen Methoden des NLP, des Textmining, der Stylometrie und Authorship Attribution sowie der Netzwerkanalyse und graphbasierten Speicherung von Informationen in Zukunft für die Geschichtswissenschaft bei Fragestellungen erwarten, die sich mit solchen Methoden beantworten lassen?⁴ Noch sind wir weder mit den Methoden noch mit den Materialien so weit, als dass diese Ansätze ihr volles Potential über die ganze Breite der historischen Forschung entfalten könnten. Vor diesem Hintergrund sind die von den Organisatoren des Forums gestellten Fragen zu beantworten, die sich zumeist schon auf Szenarien beziehen, in denen der Einsatz von Analysetools oder Statistik Ergebnisse erbracht hat, die möglicherweise neue Erkenntnisse bein-

halten oder zumindest Fragen anregen. Sicherlich führen die DH-Tools zur Erweiterung der Forschungsgegenstände und Fragehorizonte in den Geschichtswissenschaften. Die neuen Verfahren produzieren allerdings nur selten neue und unerwartete Gewissheiten. Sie provozieren hingegen regelmäßig neue Hypothesen und werden häufig zur Validierung schon bekannter Annahmen eingesetzt. In vielen Fällen wird ein durch den Einsatz eines Algorithmus errechnetes interessantes Ergebnis mehr neue Fragen als Antworten hervorbringen. Neue Forschungsbedingungen werden in der Geschichtswissenschaft zurzeit vor allem durch die digitale (und leider nur teilweise für die Benutzer kostenfreie) Bereitstellung von fachlich relevanten Informationen generiert. Prominenteste Vertreter sind hier sicher das google books-Projekt oder auch Gallica, die digitale Plattform der Französischen Nationalbibliothek. Aber auch die google ngrams sind vielleicht ein gutes Beispiel dafür, was mit solchen Textmengen möglich ist und wie gefährlich gleichzeitig eine unbedarfte Verwendung für den Wissenschaftler sein kann.⁵ Zur Vermeidung von Blindflügen müssen wir gerade in den digitalen Geschichtswissenschaften im Team zusammen mit Fachwissenschaftlern und Informatikern arbeiten. Die Zusammenarbeit auf Augenhöhe ist Voraussetzung, um anspruchsvolle Ergebnisse zu erzielen. Das wird auch den habitus des digital arbeitenden Geisteswissenschaftlers langfristig beeinflussen und die Zuschreibung von Leistungen weg von Einzelforscherpersönlichkeiten hin zu Gruppen und Laboren fördern.

Zum Schluss nun die Frage nach dem Wann. Wann kommt er denn nun, der Erkenntnisfortschritt durch Digital Humanities im oben beschriebenen Sinne? Die Antwort lautet: zum einen ist er schon da und zum anderen wird er so ganz schnell in der erträumten Breite nicht kommen, denn wir müssen noch viele Texte zeitraubend aufarbeiten, viel OCR korrigieren und viel über Sprachtechnologien und natürliche Sprache überhaupt lernen, bevor wir an der einen oder anderen Stelle mit neuen Erkenntnissen belohnt werden.



4 Für ein Experiment mit graphbasierter Datenmodellierung vgl. z. B. WETTLAUER/JOHNSON/SCHOLZ/FICHTNER/THOTEMPUDI (2013).

5 <http://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram/> (1.4.2016).

Bibliographie

- PARAVICINI, WERNER (2010), Die Wahrheit der Historiker, München (Historische Zeitschrift, Beihefte NF, Bd. 53)
- SÈRE, BÉNÉDICTE, JÖRG WETTLAUFER (eds.) (2013), Shame between punishment and penance. The social usages of shame in the Middle-Ages and Early Modern Times, Firenze [Micrologus Library 54]
- WETTLAUFER, JÖRG, CHRISTOPHER JOHNSON, MARTIN SCHOLZ, MARK FICHTNER, SREE GANESH THOTEMPUDI (2015), Semantic Blumenbach: Exploration of Text-Object Relationships with Semantic Web Technology in the History of Science, in: Digital Scholarship in the Humanities (DSH), Special Issue ›Digital Humanities 2014‹, ed. by MELISSA TERRAS, CLAIRE CLIVAZ, DEB VERHOEVEN and FREDERIC KAPLAN, Vol. 30, Supplement 1, December 2015, i187–i198
- WETTLAUFER, JÖRG, SINA WESTPHAL (2014), Digital Humanities, in: Der Archivar 67/3, 270–277