Book of Abstracts

zum

DHd Workshop

Informatik und die Digital Humanities

Leipzig, 3. November 2014

digital humanities im deutschsprachigen raum

Inhalt

Vorbemerkung ……...……. V.1

Call for Papers …… V.2

Teil A Vorträge

Kritische Geodatenanalyse – ein tragfähiger Ansatz für die Digital Humantities zwischen Geographie und Geoinformatik? ……..…..…. A.1
Tim Elrick, Dominik Kremer

Informatik und Digital Humanities: Chancen und Herausforderungen eines ungeklärten Verhältnisses …….…… A.2
Erhard Hinrichs

Differenzerfahrung und Wissensarbeit – Zwei Paradigmen zur wechselseitigen Befruchtung von Informatik und den Kulturwissenschaften ……... A.3
Reinhard Keil

The Challenge for the Computational Sciences in Digital Humanities: Establishing a Common Meta-Methodological Framework …… A.4
Jonas Kuhn

Archäo-Informatik und Digital Humanities ……..... A.5
Armin Volkmann

Teil B Diskussionsbeiträge

Blicke auf das Digitalisat in der Sammlungserschließung …… B.1
Lena Bonsiepen, Wolfgang Coy

Humanist-Computer Interaction Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik …….. B.2
Manuel Burghardt, Christian Wolff

How to integrate conservation data in the Linked Open Data Cloud …….. B.3
Ernesto William De Luca

Stemmatology, an interdisciplinary endeavor …… B.4
Armin Hoenen

Zum Stand der Interdisziplinarität in den Digital Humanities: Perspektiven aus dem Bereich der Social-Media-Forschung …… B.5
Katharina Kinder-Kurlanda, Katrin Weller

Abstract zum Workshop „Informatik für die Digital Humanities“ Eine Betrachtung aus dem Blickwinkel der Visualisierungsforschung …… B.6
Steffen Koch

On Covering the Gap between Computation and Humanities ……. B.7
Alexander Mehler, Andy Lücking

Informatik und Archäologie – Mein verflixtes 13. Jahr …… B.8
Hubert Mara

Informatik – technischer Hilfsdienst der Digital Humanities? ……. B.9
Michael Piotrowski

Auf dem Weg zu einer engeren Verzahnung von Informatik und Geisteswissenschaften ……..... B.10
Caroline Sporleder, Ralph Bergmann, Thomas Burch, Stephan Diehl, Vera Hildenbrandt, Claudine Moulin, Ingo J. Timm

Teil C Beiträge der Organisatoren

Anmerkungen zum Verhältnis von Digital Humanities und Informatik ……...……..……. C.1
Günther Görz

Informatik und die Humanities C.2
Andreas Henrich

Digital und Computational Humanities ……..…… C.3
Gerhard Heyer, Daniel Isemann

What distinguishes Humanities’ texts from other data? ……..…. C.4
Manfred Thaller

Vorbemerkung

Die vorliegende Sammlung von Kurzpapieren und Abstrakten dient der Dokumentation des vom Verband DHd getragenen und von der Universität Leipzig organisierten und ausgerichteten Workshops "Informatik und die Digital Humanities", der am 3. November 2014 in Leipzig unter Beteiligung von etwa 40 Wissenschaftlerinnen und Wissenschaftlern verschiedener Fachrichtungen stattgefunden hat. Die Auswahlkommision für den Workshop bildeten folgende Professoren:

Günther Görz, Friedrich-Alexander Universität Erlangen-Nürnberg,
Andreas Henrich, Otto-Friedrich-Universität Bamberg,
Gerhard Heyer, Universität Leipzig,
Manfred Thaller, Universität zu Köln,
Martin Warnke, Leuphana Universität Lüneburg.

Der "Call for Papers" für den Workshop wurde über einschlägige Mailinglisten der Digital und Computational Humanites Community verbreitet und die für den Workshop eingereichten Beiträge in einem mehrstufigen Auswahlverfahren von den Mitgliedern der Auswahlkommision in drei Kategorien eingeteilt: solche die zu einem Einzelvortrag eingeladen worden sind (fünf Beiträge, die in der vorliegenden Sammlung in Teil A enthalten sind), solche die nicht einzeln vorgestellt wurden, aber im Rahmen einer Diskussionsveranstaltung am Nachmittag Eingang in den Workshop gefunden haben (zehn Beiträge, die in Teil B enthalten sind) und schließlich solche welche von ihrer inhaltlichen Ausrichtung her zu weit von den intendierten Themenschwerpunkten des Workshops entfernt und folglich nicht repräsentiert waren. Die Diskussionsveranstaltung am Nachmittag ist in Form eines sogenannten Knowledge-Cafés abgehalten worden, wobei die Teilnehmer nach dem Rotationsprinzip vier verschiende Stationen mit je verschiedenen Themenschwerpunkten durchlaufen haben die jeweils von einem Mitglied der Auswahlkommision betreut und moderiert worden sind.
Als dritten und letzten Teil C enthält die vorliegende Sammlung Beiträge die die Position der Mitglieder der Auswahlkommision und Organisatoren des Workshops wiederspiegeln.
Alle Beiträge wurden auf Grundlage der eingesendeten Abstrakte ohne strikte Formatierungsvorgaben in diese Sammlung aufgenommen, was abweichende Formate und Layouts erklärt.

Im Auftrag der Auswahlkommision

Daniel Isemann
Universität Leipzig

Call for Papers

Workshop "Informatik und die Digital Humanities"

Leipzig, 3. November 2014

In den vergangenen Jahren hat das Thema "Digital Humanities" in Deutschland deutlich an Interesse gewonnen, dafür sind unter anderem auch die entsprechenden Ausschreibungen des BMBF verantwortlich. Aus den Geisteswissenschaften heraus hat sich dazu 2013 der Verband DHd ("Digital Humanities im deutschsprachigen Raum") gegründet, der im März 2014 seine erste Jahrestagung in Passau mit ca. 350 Teilnehmerinnen und Teilnehmern abgehalten hat.

Zahlreiche Informatikerinnen und Informatiker arbeiten in DH-Projekten und kooperieren mit ihren Kolleginnen und Kollegen aus den Geisteswissenschaften. Während sich Letztere nun organisieren und ein klareres Bild dieser Forschungs-richtung und -methodik zeichnen, bleibt die Perspektive der Informatik noch unscharf.

Der Workshop bedenkt deshalb das Verhältnis der Informatik zu den Geistes- und Kulturwissenschaften im deutschsprachigen Raum aus der Sicht der Informatik. Die Tagung in Passau hat gezeigt, dass es weiterer Klärung bedarf, in welcher Wechselbeziehung die Informatik zu den Fächern der Humanities steht. Kann die Kommunikation der (mindestens) "Zwei Kulturen" (C. P. Snow) erleichtert werden? Welcher Art kann ein Wissen sein, das in Mensch-Maschine-Verbünden entsteht?

Weder kann die Informatik die Probleme der Anderen schlicht lösen, noch können die Humanities hoffen oder erwarten, dass industrielle Computertechnik so ohne Weiteres eine große Hilfe bei ihren nur schwer formalisierbaren Fragestellungen sein könnte. Was ist das Besondere an den informatischen Problemstellungen der DH? Sind sie einfach nur ein weiteres Anwendungsgebiet der Informatik? Welche Teile der hermeneutisch orientierten Geisteswissenschaften informatischen Verfahren der Berechenbarkeit unterworfen werden können, wo also Computer besonders nützlich wären, das ist noch nicht restlos geklärt. Wie würde die Informatik solche Felder charakterisieren, scheint eine Typologie auf?

Entsteht da vielleicht eine neue Bindestrich-Informatik, oder sickert so viel informatische Fachkenntnis in die Geisteswissenschaften, dass diese sich selbständig machen kann? Bleibt das Abenteuer der DH also dauerhaft transdisziplinär, oder kommt es absehbar an sein Ende?

Weiterhin ist völlig offen, wie die Informatik sich selbst verändert, wenn geistes-wissenschaftliche Ansätze ins Spiel kommen, wie sie selbst davon profitiert, dass ihr Methodenspektrum und ihre Anwend-ungsbereiche sich ausweiten. Welches ist das Forschungsinteresse der Informatik an den DH? Ist es vielleicht an der Zeit, eine DH-GI-Fachgruppe zu gründen?

Der Workshop wird im Kreise der Autorinnen und Autoren der angenommenen Papers stattfinden. Alle diese Papers werden vor der Veranstaltung allen Teilnehmerinnen und Teilnehmern zugestellt, sie werden wechselseitig vorgestellt und dann diskutiert. Aus diesen Beiträgen wird auf der DH 2015 im Februar 2015 in Graz ein Panel zusammengestellt werden.

Bitte senden Sie einen Abstract von ein bis zwei Seiten Länge bis zum 31. August 2014 an:

informatik-dh-workshop2014@e-humanities.net

oder

Prof. Dr. Gerhard Heyer
Institut für Informatik
Augustusplatz 10
04109 Leipzig

Sie werden über die Annahme Ihres Beitrags bis zum 15. September 2014 unterrichtet.

Auswahlkommission
Andreas Henrich, Bamberg
Gerhard Heyer, Leipzig
Günther Görz, Erlangen-Nürnberg
Manfred Thaller, Köln
Martin Warnke, Lüneburg

Teil A

Vorträge

A.1

Kritische Geodatenanalyse – ein tragfähiger Ansatz für die Digital Humantities zwischen Geographie und Geoinformatik?

Tim Elrick – Institut für Geographie, Uni Erlangen-Nürnberg
Dominik Kremer – Institut für Kulturinformatik, Uni Bamberg

Abstract für CfP Workshop „Informatik und die Digital Humanities“, Leipzig, 3.11.2014

Kritische Geodatenanalyse – ein tragfähiger Ansatz für die Digital Humantities zwischen Geographie und Geoinformatik?

Elrick, Tim (Institut für Geographie, Uni Erlangen-Nürnberg, tim.elrick@fau.de)
Kremer, Dominik (Institut für Kulturinformatik, Uni Bamberg, dominik.kremer@uni-bamberg.de)

Nicht erst seit dem Schlagwort „big spatial data“ (Kitchin 2014, Crampton et al. 2013, Graham/Shelton 2013) erfreut sich die Analyse von georeferenzierten Daten (Geodaten) steigender Beliebtheit: immer mehr Daten unterschiedlichster Herkunft sind mit einer Georeferenzierung versehen. Dies macht das Feld der *Digital Humanities* auch für die Humangeographie und die Geoinformatik interessant[1].

Einige Forschungstrends in beiden Disziplinen, der Humangeographie und der Geoinformatik, weisen derzeit inhaltlich erstaunliche Schnittmengen auf: (1) Anstelle von statischen Strukturmodellen ist eine räumliche (z.B. Mobilität) wie auch zeitliche (z.B. ereignisbasierte) Dynamisierung bei der Modellierung von Prozessen zu beobachten (Cresswell 2006 für die Humangeographie, Grenon/Smith 2004 für die Geoinformatik). (2) Ähnlich zum hermeneutischen Grundverständnis interpretativ arbeitender Ansätze der qualitativen Textanalyse (z.B. Glaser/Strauss 2009 für die Humangeographie) steht auch bei der Ausgestaltung von Mensch-Maschine-Schnittstellen beim Sprechen über Räume zunehmend die Operationalisierung als individueller Aushandlungsprozess zwischen dem Sender und dem Empfänger einer Nachricht im Vordergrund (z.B. Winter/Freksa 2014 für die Geoinformatik). (3) Verstehen und Handeln wird dabei immer seltener als rein kognitive (intentionale) Leistung verstanden, sondern als hoch-routiniertes Verhalten in Abhängigkeit von der jeweiligen Sozialisierung (Schatzki 2002 für die Humangeographie, Davies/Pederson 2001 für die Geoinformatik).

Dabei stellt sich die Frage, welche Wechselbeziehung die beiden genannten Disziplinen eingehen (können); inwiefern sie produktiv zusammenarbeiten können. Dieser Frage soll anhand des Beispiels der kritischen Geodatenanalyse nachgegangen werden: Anhand der Forschungen zu OpenStreetMap, der derzeit größten kollaborativ erzeugten Geodatenbank, wollen wir den Ansatz der *Kritischen Geodatenanalyse* vorstellen und die Wechselbeziehungen zwischen informationstechnischer und geistes-/sozialwissenschaftlicher Herangehensweise diskutieren. Während in der Geoinformatik Qualität und Abdeckung (z.B. Keßler/de Groot 2013) gemessen und Aktivitätsmuster der Nutzer (z.B. Neis/Zipf 2012) extrahiert werden, fokussieren humangeographische Arbeiten im Umfeld der Kritischen Kartographie (Glasze 2009) auf die gesellschaftlichen Implikationen der zur Verarbeitung der Geodaten genutzten technischen Verfahren (Bittner 2014). Im Schnittfeld zwischen formaler Repräsentation und inhaltlicher Interpretation fällt so z.B. auf, dass die Einstiegshürden – anders als in der klassischen Kartographie – zwar nicht mehr autoritativer Art (z.B. Behörden), dafür aber technischer Art sind, bei der nur noch wenige Nutzer einer digitalen Elite den Institutionalisierungsprozess überhaupt steuern können (Graham et al. 2013). Kitchin et al. (2013) fordern daher einen Perspektivwechsel von der

[1] *Digital Humanities* wird hier verstanden als weiter Begriff, der nicht nur geisteswissenschaftliche, sondern auch sozialwissenschaftliche informationstechnisch-gestützte Forschung beinhaltet.

Karte hin zu einem Organisationsprozess von Geodaten. Während geoinformatische Analysesoftware die funktionalen Abhängigkeiten in der Software beschreiben und die strukturellen Regularitäten bei ihrer Verwendung offen legen kann (z.B. Kremer/Stein 2014), bieten humangeographische Studien ein Analyseframework (Bittner et al. 2013), das das Miteinbeziehen technischer Verarbeitungszusammenhänge für eine Analyse der inhaltlichen Aussagekraft der Geodaten (z.B. Elrick 2014) mit einbezieht.

Aus diesem interdisziplinären Forschungsbeispiel möchten wir verallgemeinerbare Herausforderungen und Lösungsansätze bei Forschungsprojekten im Schnittbereich zwischen Informatik und Kulturwissenschaften ableiten.

Bittner, C. (2014). Reproduktion sozialräumlicher Differenzierungen in OpenStreetMap: das Beispiel Jerusalems. *Kartographische Nachrichten*, *64*(3), 136–144.

Bittner, C., Glasze, G., & Turk, C. (2013). Tracing contingencies: analyzing the political in assemblages of web 2.0 cartographies. *GeoJournal, 78*(6), 935–948.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, *40*(2), 130–139.

Cresswell, T. (2006). *On the move: mobility in the modern western world*. Taylor & Francis.

Davies, C., & Pederson, E. (2001). Grid patterns and cultural expectations in urban wayfinding. *Spatial Information Theory* (pp. 400–414). Springer.

Elrick, Tim (2014). Sozialwissenschaftliche *tag*-Analyse mit OpenStreetMap-Daten am Beispiel religiöser Andachtsstätten in Deutschland. *Kartographische Nachrichten 64*(3), 152–156.

Glaser, B. G., & Strauss, A. L. (2009). *The discovery of grounded theory: Strategies for qualitative research.* Transaction Publishers.

Glasze, G. (2009). Kritische Kartographie. *Geographische Zeitschrift*, 181–191.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, *3*(3), 255–261.

Graham, M., Zook, M., & Boulton, A. (2013). Augmented reality in urban places: contested content and the duplicity of code. *Transactions of the Institute of British Geographers*, *38*(3), 464–479.

Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial cognition and computation, 4*(1), 69–104.

Keßler, C., & de Groot, R. T. A. (2013). Trust as a proxy measure for the quality of Volunteered Geographic Information in the case of OpenStreetMap. *Geographic Information Science at the Heart of Europe* (pp. 21–37). Springer International Publishing.

Kitchin, R., Gleeson, J., & Dodge, M. (2013). Unfolding mapping practices: A new epistemology for cartography. *Transactions of the Institute of British Geographers, 38*(3), 480–496.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12.

Kremer, D. & Stein K. (2014). Ein Analyseansatz für Nutzerverhalten auf Basis von OSM-Daten. *Kartographische Nachrichten 64*(3), 144–152.

Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information, 1*(2), 146–165.

Schatzki, T. (2002). *The site of the social: A philosophical exploration of the constitution of social life and change.* Pennsylvania State University Press.

Winter, S., & Freksa, C. (2014). Approaching the notion of place by contrast. *Journal of Spatial Information Science*, (5), 31–50.

A.2

Informatik und Digital Humanities: Chancen und Herausforderungen eines ungeklärten Verhältnisses

Erhard Hinrichs – Allgemeine Sprachwissenschaft und Computerlinguistik, Seminar für Sprachwissenschaft, Universität Tübingen

Informatik und Digital Humanities: Chancen und Herausforderungen eines ungeklärten Verhältnisses

Autor: Erhard Hinrichs, Allgemeine Sprachwissenschaft und Computerlinguistik, Seminar für Sprachwissenschaft, Universität Tübingen

Abstract eingereicht zum Workshop "Informatik und die Digital Humanities"; Leipzig, 3. November 2014

Ziel des Workshops "Informatik und die Digital Humanities“ ist es, das Verhältnis der Informatik zu den Geistes- und Kulturwissenschaften im deutschsprachigen Raum aus der Sicht der Informatik zu bedenken. Der Call for Papers weist mit Recht darauf hin, dass die Digital Humanities sich „... nun organisieren und ein klareres Bild dieser Forschungsrichtung und -methodik zeichnen“. Dennoch ist festzuhalten, dass es sich bei den Digital Humanities sowohl hinsichtlich ihres Forschungsgegenstandes als auch hinsichtlich ihrer Methodik nach wie vor um ein „moving target“ handelt, auch wenn die Umrisse des Zielbereichs sich nun klarer erkennen lassen. Die Heterogenität des Begriffs Digital Humanities rührt nicht zuletzt von den Geisteswissenschaften und Kulturwissenschaften selbst her, die sich durch eine große methodische Vielfalt und durch unterschiedliche Forschungs“gegenstände“ wie Texte, Objekte und Ereignisse auszeichnen. Die Frage nach dem Verhältnis der Informatik zu **den** Geistes- und Kulturwissenschaften fordert also eine Gegenfrage oder eine differenzierende Begriffsklärung bewusst oder unbewusst geradezu heraus.

Führende Vertreter der Digital Humanities sind sich der Schwierigkeiten bei der Begriffsdefinition der Digital Humanities nur allzu bewusst. In seiner Rede anlässlich der Auszeichnung mit dem Zampolli Preis auf der *Digital Humanities 2014* Tagung, die er im Juli 2014 in Lausanne gehalten hat, betont Ray Siemens, Canada Research Chair in Humanities Computing und Distinguished Professor in the Faculty of Humanities an der University of Victoria, daher den Charakter einer sich nach wie vor noch entwickelnden Forschungspraxis und die Wichtigkeit des „Community Buildings“ und der Förderung des wissenschaftlichen Nachwuchses im Umgang mit innovativen Forschungsmethoden gegenüber Versuchen, die Digital Humanities als Kampfbegriff zu benutzen, um einzelne geisteswissenschaftliche Disziplinen und Methoden auszugrenzen, einer Tendenz, vor der auch McCarty (2006) eindringlich warnt.

Von manchen WissenschaftlerInnen sowohl innerhalb als auch außerhalb der Geistes- und Kulturwissenschaft wird die Sinnhaftigkeit des Begriffs der Digital Humanities als solcher angezweifelt mit dem Hinweis, dass er bestenfalls den Einsatz von computergestützten Verfahren in der geistes- und kulturwissenschaftlichen Forschung benennen könne, somit aber keinen eigenen Gegenstandsbereich ausmachen könne und dürfe. Schließlich gäbe es ja auch keine Digital Chemistry oder Digital Economics. Schließt man sich dieser Sichtweise an, so kommt der Informatik für die (digitalen) Geisteswissenschaften, primär die Rolle einer Hilfswissenschaft zu, um die eigentliche geisteswissenschaftliche Forschung mit Techniken und Methoden der Informatik zu befördern – eine für InformatikerInnen sicherlich nicht sonderlich attraktive Rolle und kein hinreichender Grund, eine neue „Bindestrich“-Informatik zu konstituieren.

Eine tragfähigere, weil reizvollere, Brücke für die Zusammenarbeit zwischen Informatik und den Geistes- und Kulturwissenschaften bietet Willard McCarty, einer der Pioniere der Digital Humanities und Empfänger des Roberto Busa Preises (2013), mit dem Begriff der Modellierung als Kernbegriff für die Digital Humanities an: „The question of modeling arises naturally for humanities computing from the prior question of what its practitioners across the disciplines have in common. [...] I argue for modeling as a model of such a practice.“ (McCarty 2008). Dass der Begriff der Modellierung in sich selbst ein schillernder Begriff ist, der von verschiedenen FachvertreterInnen innerhalb der Digital Humanities durchaus

verschieden und kontrovers interpretiert wird, darauf hat die Panel Session zum Begriff der Modelling auf der Digital Humanities 2014 in Lausanne, eindrucksvoll aufmerksam gemacht.

Trotz der skizzierten Schwierigkeiten bietet der Begriff der Modellierung die Chance einer Übertragung und eines Dialogs mit der Informatik, für die der Begriff der komputationellen Modellierung zentrale Bedeutung hat. Dies betrifft sowohl die angewandte Informatik mit Methoden der Datenbanktechnologie, des Data Mining, der Mustererkennung und der Visualisierung, die ohnehin schon wichtige Forschungs- und Anwendungsfelder für die Informatik und für die Digital Humanities gleichermaßen darstellen und bei der die geisteswissenschaftlichen Forschungsgegenstände mit semi-strukturierten und Daten eine wichtige und reizvolle Herausforderung darstellen. Der Begriff der Modellierung bietet aber auch eine sinnvolle Brücke zu Grundlagenfragen der Informatik, etwa nach dem Begriff der Komputation als solcher, dessen Verhältnis zum menschlichen Denken und Verstehen mindestens klärungswürdig ist. Er verweist zugleich über die Geisteswissenschaften hinaus auf den Zusammenhang mit den Neurowissenschaften und auf das Neurocomputing hin, der mindestens für die Philosophie des Geistes zu einer innovativen Forschungsrichtung geführt hat.

Ungeachtet der dabei zu erzielenden neuen Erkenntnisse, bietet der Begriff der Modellierung die Chance, Methoden der Informatik und der (digitalen) Geisteswissenschaften gegenüberzustellen, um Gemeinsamkeiten und Unterschiede zwischen den damit benannten Einzeldisziplinen auszuloten.

Schließlich bietet der Begriff der Modellierung auch einen Hinweis auf eine mögliche neue Bindestrich-Informatik. Hier sei vergleichend auf die Konstituierung der Wirtschaftsinformatik verwiesen, für die ein ebenfalls hinreichend abstrakter Begriff der Modellierung von Informations- und Kommunikationsprozessen Pate gestanden hat.

Ob eine neue Bindestrichinformatik sinnvoll und zukunftweisend sein kann, wird vermutlich erst die Zukunft zeigen. Schon jetzt ist zu konstatieren, dass die wissenschaftliche Praxis in diese Richtung weist. So hat Stanford University, die als hervorragend ausgestattete und international renommierte Privatuniversität nicht darauf angewiesen ist, ihr Lehr- und Forschungsportfolio opportunistisch auszurichten, vor kurzem zwei neue interdisziplinäre Studiengänge auf der Schnittstelle von Geisteswissenschaften und Informatik aus der Taufe gehoben und sich damit an internationale Trends mit ähnlichen Bemühungen angeschossen; vgl. http://news.stanford.edu/news/2014/march/facsen-joint-majors-030614.html.

Literaturangaben:

Willard McCarty (2006). Tree, Turf, Centre, Archipelago—or Wild Acre? Metaphors and Stories for Humanities Computing. *Literary and Linguistic Computing*, Vol. 21.1, 1-13.

Willard McCarty (2008). Modelling: a Study in Words and Meanings, In: S. Schreibman et al. (Eds). *A Companion to Digital Humanities*, Wiley Publishers.

A.3

Differenzerfahrung und Wissensarbeit – Zwei Paradigmen zur wechselseitigen Befruchtung von Informatik und den Kulturwissenschaften

Reinhard Keil – Heinz Nixdorf Institut, Universität Paderborn

Differenzerfahrung und Wissensarbeit –

Zwei Paradigmen zur wechselseitigen Befruchtung von Informatik und den Kulturwissenschaften

Reinhard Keil
Heinz Nixdorf Institut, Universität Paderborn

Reinhard.Keil@uni-paderborn.de

Vorbemerkung

Die nachfolgenden Überlegungen basieren auf Erfahrungen bei der Durchführung von zwei Projekten im Bereich der Digital Humanities und der Beteiligung an der (erfolgreichen) Antragstellung eines Zentrums für Digital Humanities an der Universität Paderborn, das seine Arbeit zum 1. September aufnimmt.

Kurzfassung

Traditionell tun sich die Geisteswissenschaften und die Informatik noch schwer in der Zusammenarbeit, scheinen doch die beiden Wissenschafts- und Wissenskulturen nicht nur unverbunden nebeneinander zu stehen, sondern sich in Teilen sogar zu widersprechen – zumindest dann, wenn es um die formale Modellierung von Wissen geht (vgl. Osterloh 2006), tauchen Begriffe wie hermeneutische Auslegung, Interpretation und Deutung auf, die einer Formalisierung grundsätzlich zu widersprechen scheinen.

Dies ändert sich jedoch, wenn man nicht mehr davon ausgeht, Wissen vollständig formal modellieren zu wollen, sondern teilweise auf algorithmischen Ableitungen basierende Ergebnisse nutzt, um neue Einsichten über den Gegenstandsbereich zu gewinnen (Differenzerfahrung). Ähnlich der wissenschaftlichen Visualisierung ist es mit vielen Informatikmethoden möglich, strukturelle Zusammenhänge offenzulegen, ohne zugleich den Anspruch zu erheben, den Gegenstand selbst erschöpfend modelliert zu haben. Statistische Analysen großer Textcorpora (Big Data) gehören z.B. ebenso zu diesem Methodenrepertoire wie Ontologien (wobei schon der Plural eine informatikspezifische Sicht verdeutlicht). Stellt man das Paradigma der Differenzerfahrung in den Vordergrund, geht es nicht mehr darum, wer die Deutungshoheit über einen Gegenstandsbereich innehat oder ob dieser überhaupt einer formalen Behandlung zugänglich ist, sondern mit welchen Teilstrukturen, Differenzierungen und alternativen Gegenüberstellungen neue Einsichten über den Gegenstandsbereich gewonnen werden können (vgl. Keil 2010 sowie Schulte et al. 2011). Da es hier um das Wechselspiel zwischen Verstehen und Modellieren geht, ist ein interdisziplinärer Zugang unerlässlich.

Das Paradigma der Differenzerfahrung dient ebenso wie das Paradigma der Wissensarbeit (siehe Drucker 1957) der Unterstützung der Forschungsarbeit und des (interdisziplinären) Forschungsdiskurses. Hier geht es u.a. darum, ein Gestaltungsrepertoire bereit zu stellen, das es gestattet, kooperative und verteilte Forschungsarbeiten zu unterstützen, indem Wissensartefakte (vgl. das Konzept der Boundary Objects in Starr, Griesemer 1989) unter Bezug auf vorhandene Repositorien Gegenstand der Manipulation und des Diskurses sind. Dieses Wissen muss über teilweise große Zeiträume hinweg erschlossen, tradiert, über verteilte Standorte hinweg zusammengeführt und wieder verteilt werden. Dabei gilt es, bedeutsame Verknüpfungen zwischen Medienobjekten zu erkennen, die in dieser Form nicht explizit angelegt und physisch repräsentiert sind.

Dabei tritt eine Fülle von Fragen auf, die sowohl für die Modellierung auf Seiten der Informatik als auch für die Analyse und Handhabung auf Seiten der Geisteswissenschaften, denn mit digitalen Medien ist es möglich, das Objekt der Wahrnehmung zugleich zum Objekt des Handelns und der Manipulation zu machen. Was sind beispielsweise bei einer musikalischen Notation die einzelnen Objekte? Einzelne Noten, Takte, etc. Und wie sieht die Reichweite der Zeichen aus, d.h. auf welchen Bereich der nach folgenden Zeichen beziehen sie sich? Will man hier unterschiedliche Interpretationen zulassen, muss man einen Weg finden, diese auch in der digitalen Welt zu repräsentieren. Damit tritt erneut eine Fülle von Fragen auf, die von der Handhabbarkeit über die Modellierung bis hin zur Entwicklung von Werkzeugen reicht, mit denen z.B. textuelle und nicht-textuelle Medienobjekte unterschiedlicher Granularität verknüpft und adäquat gehandhabt werden können.

Anhand von Beispielen aus den drei oben angegebenen Projektbereichen sollen die beiden hier skizzierten Paradigmen vorgestellt und der Nutzen einer interdisziplinären Zusammenarbeit zwischen der Informatik und den Kulturwissenschaften verdeutlicht werden.

Literatur

Drucker, P.F. (1957): The Landmarks of Tomorrow. New York, Harper & Row

Keil, R.: Das Differenztheater. Koaktive Wissensarbeit als soziale Selbstorganisation. In: Bublitz, H., Marek, R., Steinmann, C., Winkler, H. (Hrsg.): Automatismen. Wilhelm Fink: München, 2010, S. 205-229

Osterloh (2006): Human Resources Management and Knowledge Creation, in: Nonaka, I./ Kazuo, I. (Hrsg.) *Handbook of Knowledge Creation*, Oxford, Oxford University Press, S. 158-175.

Schulte, J.; Keil, R. & Oberhoff, A.: Unterstützung des ko-aktiven Forschungsdiskurses durch Synergien zwischen E-Learning und E-Science. In: Köhler, Th., Neumann, J. (Hrsg.): Wissensgemeinschaften: Digitale Medien - Öffnung und Offenheit in Forschung und Lehre. Münster: Waxmann, 2011, S. 81-91

Star, S. L. & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professional in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*, 387–420.

A.4

The Challenge for the Computational Sciences in Digital Humanities: Establishing a Common Meta-Methodological Framework

Jonas Kuhn – University of Stuttgart

The Challenge for the Computational Sciences in Digital Humanities: Establishing a Common Meta-Methodological Framework

Jonas Kuhn, University of Stuttgart (*jonas.kuhn@ims.uni-stuttgart.de*)
August 2014

In my view, *the* major challenge for the branch of Digital Humanities that aims to exploit the potential in scalable computational models is to go beyond the status quo in DH where the relatively few studies going to **greater analytical depth** typically involve custom-tailored technological solutions which are developed and implemented in direct tandem collaborations and whose portability to other studies/disciplines is often unclear. Non-trivial solutions typically have to draw upon a complex collection of resources that require an appropriate adjustment and adequate combination techniques. Very often, the working assumptions of certain tool components will fail to be met in a different domain, which can cause unexpected and hard to predict system behavior even when formally, all interfaces adhere to the required exchange formats. The **lack of portability and scalability (at moderate costs)** is a serious issue, since it can lead to disappointment at various points:

- *Engineering efficiency*. Funding sources investing money as well as humanities partners investing considerable time in interface specifications, guidelines and (test) annotation will expect that future projects can re-use major parts of earlier efforts
- *Theoretical flexibility*. Scholars applying computational modeling frameworks will hope to be able to factorize out theory-specific assumptions from a prior empirical study while still taking advantage of the algorithmic facilities
- *Scientific rigor*. Ill-informed re-use of modeling solutions outside of the intended scope is likely to lead to unnoticed artifacts in the system predictions; thus there is a considerable risk that apparent break-throughs from DH methods will not stand the test of critical assessment by other scholars from the disciplines or will turn out impossible to replicate. This may discredit the use of such techniques.

One way to respond to these issues is to restrict DH support to generic exploration tools that avoid going into greater analytical depths. But this would sacrifice most of the goal of opening up innovative avenues of research in the digital humanities (besides the fact that it is probably an illusion that there are safe generic analytical steps, given the diversity of data resources): generic support tools often lead to scholars' disappointment, since the effort of including them in one's working methodology does not pay off. There is so much manual work left that it seems more efficient to leave out the analytical support step completely.

I am convinced that effective use of computational models in the humanities can be a success. There is great potential in the modeling frameworks that have been developed over the last decades (for instance in computational linguistics and language technology). But to exploit them, the field needs a more **systematic methodology** that breaks down analytical processes into building blocks whose **"deeper" functionality** is **transparent to the users** in the humanities, so they are in a position to make their own critical assessment of the reliability of a particular component or component chain — and arrange for adjustments as necessary. Crucially, the meta-architecture to be established should include best practices for non-computational intermediate steps too, which are required to bridge the methodological gap between data-based empirical results and higher-level disciplinary research questions. Ultimately, digital humanities scholars should feel fully competent to draw upon a flexible methodological toolbox so they can try backing up any partial results from one component with evidence obtained from other sources, make informed adjustments to the components, or attempt an entirely different way of approaching the available information sources.

In other words, the mid- to long-term goal should *not* have IT specialists optimize a tool chain for fully automatic analysis so as to achieve the best possible performance for some specified task (which is bound to be imperfect for any non-trivial question anyway, thus requiring a responsible integration into higher-level research questions). The digital humanities should rather aim to create transparency within a complex multi-purpose system of interacting information sources of variable quality or reliability (in plain extension of the classical competences humanities scholars have always had regarding approaches to their object of study). Contrary to the assumptions one can make about the typical users in a standard web-oriented application scenario of language technology and visual analytics (where users rarely have any philological or other meta-level attachment to the text basis from which they are seeking information), humanities scholars have far-reaching competences and intuitions about their objects of study and their sources. This makes the goal of developing an interactive framework for a network of knowledge sources a promising endeavor, drawing on techniques for aggregation, diagnostic and

explorative visualization, quantitative analysis and linking back to data instances and (re-)annotation tools etc. In particular this means that C.S. is not in the role of a mere service provider, but the challenges that the use of **computational models in the humanities** pose are **genuine scientific challenges for Computer Science (or Informatics)** in its own right.

My hope is that a comprehensive and **truly interdisciplinary digital humanities methodology** will provide an effective choice of flexibly adjustable tools which are accessible to critical reflection and compatible with classical disciplinary methods and will thus lead to a break-through **overcoming** the **residual wide-spread limitations** on the types of higher-level subject-specific research questions that are being actively addressed in DH work — despite the availability of growing amounts of digitized resources.

The most critical issue in practice, I believe, will be to **trade off** the mentioned **optimization criteria** of *Engineering efficiency*, *Theoretical flexibility* and *Scientific rigor*. For establishing innovative computational methods in multiple disciplines it is important for the partners in the computational sciences to showcase some effective ways of porting and scaling up solutions without extremely high costs and effort. Otherwise the established cores of the humanities disciplines will not see the merit of a paradigm shift — so DH would remain a special community that is not anchored in the relevant fields. During some transitional period, it may hence be strategically justified to place slightly less emphasis on a rigorous checking of all prerequisite assumptions — which might slow down imaginative explorations of newly established approaches. However, this means walking a thin line. Since the scholar is losing direct control over the data when using computational models, it is crucial to establish standardized evaluation techniques. Otherwise, artifacts in computational modeling results are highly likely to give rise to misinterpretations, which would discredit the approaches in the core disciplines. In the long run, the digital humanities should place utmost importance on highest scientific standards in their methodology to avoid any impression that computational models can be used to bring about effects that are not warranted empirically.

So far, the main emphasis in this short contribution has been on the viewpoint from the computational sciences asking in which ways they may have to adjust to the situation to facilitate effective interdisciplinary work. Of course the success also depends on some degree of **flexibility on the side of the humanities partners**. From own and reported experience, the most critical obstacle to establishing an effective collaboration is the **fundamentally different role** that a **meta-level understanding of methodological challenges** plays on the two sides: for computational scientists, the ultimate driving force for scholarly work that doesn't even need to be mentioned is finding a general solution to a *class* of similar problems. Hence, the expectation is that it is required and helpful to switch back and forth between problems that are related/similar along certain dimensions. For the humanist on the other hand, the connection to other work conceived of as most important is often at a level of disciplinary content and theory that is inaccessible to a computational scientist, whereas the methodologically most similar DH work is in a more distant content domain or sub-discipline that the humanist hesitates to relate to. It is very important for interdisciplinary success to overcome this dilemma: the solution may lie in an even more multi-disciplinary build-up of research teams and **the involvement of "mediators",** i.e., methodologically aware DH scholars that have a high-level background in a number of disciplinary areas and can relay content questions to computational scientists and methodological constraints to humanist specialists.

As a side note, I believe that the field of **digital humanities** can **learn** a lot from the **experience in the field of computational linguistics**, which has faced many of the technical and methodological issues over three decades. In the many strands of research in computational linguistics, formal linguistics, corpus linguistics and data-driven language technology, various different methodological ways of bringing together apparently incompatible modeling assumptions have been explored to quite some depth (and in particular the issues of empirical evaluation methods have been debated for a long time). (This is not to say that computationally ambitious branch in digital humanities should or could simply take over solutions one-to-one from computational linguistics; there may be mistakes that the digital humanities can avoid, and the broader spectrum of research questions and objects of study addressed obviously places additional challenges. But it would be rather unreasonable not to take advantage of the many years of practical experiments in sociology of science under similar conditions…)

A.5

Archäo-Informatik und Digital Humanities

Armin Volkmann – Cluster of Excellence "Asia and Europe in a Global Context", Heidelberg University

Archäo-Informatik und Digital Humanities

In den archäologischen Wissenschaften sind die Anwendungen von informations-technologischen Methoden im Vergleich zu anderen geisteswissenschaftlichen Disziplinen weit verbreitet und seit einigen Jahren Standard im Dokumentationsverfahren der Ausgrabungen. In der Auswertung der Grabungsergebnisse ist deren Einsatz oft durch das Wissen und Können des jeweiligen Anwenders beschränkt. So tritt die Archäologie meist nur als „Sekundärverwerter“ von vielerlei Tools und Methoden auf. Beispielsweise sei hier der konsequente Einsatz von Geografischen Informationssystemen (GIS) in den Fundstellenarchiven genannt, der die Landschaftsarchäologie zu siedlungsarchäologischen Fragestellungen seit den 1990er Jahren in großen Schritten vorwärts brachte und noch immer bringt. Weitere Innovationsschübe wurden durch den Einsatz von 3-D-Technologien erreicht, die ebenso in den archäologischen Fächern nun weit verbreitet sind. Insgesamt hat die Digitalisierung der Archäologie einen Prozess eingeleitet, der nicht mehr umkehrbar erscheint und neben der Forschung auch in der Lehre in spezifischen Lehrveranstaltungen (z.B. zu 3D, GIS, CAD und Datenbanken etc.) seinen Niederschlag findet.

Beschäftigt man sich im Detail mit informationstechnologischen Fragestellungen in der Archäologie, so werden doch vielerlei Defizite offensichtlich, die Großteils in der Sekundärverwertung der Methoden und Tools als Anleihen aus der Informatik ohne fachorientierte Anpassung begründet sind. So sind sie für den „Durchschnitts-Geisteswissenschaftler/Archäologen“ schon anwendbar, aber nur ansatzweise nachvollziehbar, sodass ein direkter Dialog zur Informatik mit „Bug-Feedback“, Verbesserungswünschen oder Methodenentwicklung nur selten erfolgt. Sowohl die Anzahl als auch die Geschwindigkeit der sich dynamisch entwickelnden Forschungen der Angewandten Informatik sind aus geisteswissenschaftlicher Perspektive mittlerweile nur sehr schwer überschaubar. Auch sind die grundlegend divergenten Forschungsansätze der hermeneutischen Archäologie und empirischen Informatik eine beiderseitige Hürde des direkten Wissenstransfers. In vielerlei gemeinsamen Projekten taucht noch immer der Gedanke der „Hilfswissenschaften“ aus der jeweiligen Perspektive auf, wobei nur im seltenen Ausnahmefall Informatik und Geisteswissenschaften gleichberechtigte Partner sind, die für beide Disziplinen innovative Ergebnisse generieren.

Gerade in Bezug dessen ist aber die Stärke eines Bindestrichfachs, wie Archäo-Informatik, begründet: So kann der Dialog und Wissenstransfer hoch effektiv vorangebracht werden; so können echte transdisziplinäre Projekte entwickelt werden, die nicht nur interdisziplinären Charakter (für ein Drittmittelprojekt) haben. Und so kann die Diskrepanz in den Entwicklungsgeschwindigkeiten und den Forschungsansätzen überwunden werden. Innovationsschübe sollten dabei wechselseitig fließen und zwar nicht durch einfache Adaption, sondern durch fachspezifische Entwicklung von Methoden und Anwendungen, die dabei auch neue Felder für die Informatik eröffnen, wie beispielsweise die Quanti- und Qualifizierung von imperfekten, „weichen“, teils nicht gemessenen Datensätzen aus den oft noch „klassisch“ Daten erhebenden Geisteswissenschaften.

Aber wie und wo verortet man ein solches Fach? An der Philosophischen oder Naturwissenschaftlich-Mathematischen Fakultät? Wie schafft man möglichst zeitnah transdisziplinäre Studien- und Promotionsordnungen, gerade hinsichtlich der Anwerbung von (internationalen) Spitzenkräften? Könnte eine thematisch-fachliche und schwerpunktmäßige Ausrichtung der gerade erst entstehenden Curricula der Digital Humanities als implizites Fach hier ein Lösungsansatz sein? Sollten sich die Digital Humanities primär auf textorientierte Forschungen konzentrieren, wie teils gefordert wird? Oder ist

das weit gefächerte Spektrum der Digital Humanities mit sehr heterogenen Datenquellen, aus Text-, Bild-, Ton-, Video- und Messdaten, wie beispielsweise aus der Archäologie, gerade die große Chance für die Entwicklung von anspruchsvollen Forschungsansätzen mit gleichberechtigten Partnern der Informatik und Geisteswissenschaft?

Die Digital Humanities sollten keine Generalisten hervorbringen, um sich deutlich von der oberflächlichen Digital Literacy abzusetzen und der Gefahr der jüngst unterstellten Trivialität entgegen zu wirken. Die Digital- oder Informationskompetenz sollte durch Module an der Informatik und Geisteswissenschaft in einem Grundstudium und einer aufbauenden Phase der thematischen Ausrichtung mit einer Spezifizierung hin zu einer Disziplin erreicht werden, wobei in der fachorientierten Anwendung informatischer Methoden in eher informatiknahen Bereichen der Geisteswissenschaften ein beiderseitig ergiebiger Wissenstransfer begründet zu sein scheint. Im Fall der Archäo-Informatik stellt z.Z. die Entwicklung von spezifischen Archäologischen Informationssystemen (AIS) mit der Implementierung von heterogenen Daten eine große Herausforderung dar, die nicht einfach durch die Anpassung von GIS zu lösen ist, da z.B. neben den raumbezogenen Daten auch nichtraumbezogene Daten, wie zur Kultur, Fundumständen usw., ein großes Gewicht aufweisen. In der kollaborativen Methodenentwicklung, u.a. zur Standardisierung imperfekter Datensätze der Angewandten Informatik bzw. Geo-Informatik mit der Archäologie, können grundlegend neue Forschungsansätze entstehen, die deutlich über noch immer bestehende Forschungsgrenzen einzelner Disziplinen mit eigenen nichtinteroperablen Datenstandards und Methoden hinausgehen.

Dr. Armin Volkmann
Junior Research Group Leader "Digital Humanities/Digital Cultural heritage"
Cluster of Excellence "Asia and Europe in a Global Context"
Heidelberg University
Karl Jaspers Centre
Voßstr. 2, Gebäude 4400
69115 Heidelberg
Tel. +49 (0)6221 54 4310
http://www.asia-europe.uni-heidelberg.de/en/research/jrg/jrg-digital-humanities.html
http://archdigi.hypotheses.org/

Teil B

Diskussionsbeiträge

B.1

Blicke auf das Digitalisat in der Sammlungserschließung

Lena Bonsiepen, Wolfgang Coy – Exzellenzcluster „Bild-Wissen-Gestaltung“ der Humboldt-Universität zu Berlin

Abstract für den Workshop „Informatik und die Digital Humanities“, Leipzig, 3. November 2014

Blicke auf das Digitalisat in der Sammlungserschließung

Lena Bonsiepen, Wolfgang Coy

Ein zentrales Thema zwischen Informatik und den geistes- und kulturwissenschaftlichen Anwendungen ist die Interpretation des Digitalisats. Mit dem Prozess der Digitalisierung mithilfe geeigneter Sensor- und Messtechnik wird entlang einer gemeinsam zu entwickelnden Modellvorstellung eine digitale Beschreibung der zu untersuchenden Objekte und Kontexte konstruiert. Diese muss einerseits den technischen Verwertungsmöglichkeiten entsprechen und zum andern eine handhabbare Struktur für die Nutzung in den Anwendungen bieten.

Anforderungen an das Digitalisat sind je nach Ausgangslage und Perspektive unterschiedlich. Aus der Sicht der Informatik ist es von zentraler Bedeutung, dass die Formalisierung den Ansprüchen an eine programmierbare Vorgehensweise, an eine hinreichend präzise Differenzierung der Messwerte, an eine angemessene Speicherung und eine langfristige Organisation der Archivierung und Bewahrung erfüllt werden. Aus der Anwendersicht werden diese technischen Fragen eher als irgendwie zu realisierende Randbedingungen gesehen; hier spielt die vielfältige Beschreibungsmöglichkeit und eine im Fortgang der Arbeit bei Bedarf revidierbare Anpassung an die inhaltlichen Fragestellungen eine wesentlich größere Rolle.

Mit der Vielzahl unterschiedlicher Sensoren hat sich das Feld des Digitalisierbaren enorm erweitert. War am Anfang die Lochkarte und der Papierausdruck das Kommunikationsmittel der Wahl für die frühe Rechentechnik, sind heute Bildaufnahme, Tonaufnahme oder Video selbstverständlicher Elemente der Digitalisierung geworden. Dies schließt längst weitere Messtechnik, etwa der Materialforschung mit UV und Infrarot, Röntgenstrahlung oder Ultraschall ein. Neben der zweidimensionalen Bildverarbeitung und der Verarbeitung von Bildfolgen werden nun auch dreidimensionale Aufnahme, Speicherung und Bearbeitungstechniken rechnergestützt integriert. Im Rahmen der Erschließung einer Gemäldesammlung werden diese unterschiedlichen Ansätze der Materialwissenschaft und Informatik mit den Fragen der Kunst- und Bildwissenschaft zusammengeführt.

Digitalisierung bleibt nicht bei der Objektbeschreibung stehen. Die Integration von Digitalisaten in den realen Kontext von Räumen, Gebäuden und mobilen Umgebungen bildet eine neue Herausforderung für das Verhältnis von realem zu virtuellem Raum. Ebenso können Arbeitsvorgänge und weitere Prozesse digital abgebildet und unterstützt werden. Im informatischen Kern stehen dabei Modellierung und digitale Simulation, die mit diesen Datenströmen virtuelle Systeme und Räume aufbauen. Dazu werden Sensor- und Aktortechniken an Digitalspeicher, Rechner und Rechnernetze programmiert angebunden.

Die Erstellung, die Verwaltung und der Umgang mit Digitalisaten und virtuellen Räumen formt die Arbeits- und Lebenswelten des 21. Jahrhunderts. Unser alltägliches, kulturelles und

gesellschaftliches Verständnis unserer Umwelt wird damit grundsätzlich in Frage gestellt und herausgefordert. Stichworte wie „vernetzter Wissenschaftlerarbeitsplatz“, „Digital Humanities“ oder „Industrie 4.0“ benennen Aspekte dieser Veränderung, die in ihrer Gesamtheit freilich weit über die einzelnen Sichten hinausgehen.

Wir können die genannten Aspekte aus unserer Praxis der Erschließung einer Gemälde- und Objektsammlung in Zusammenarbeit mit der Kunstbibliothek Berlin belegen und genauer erläutern.

Berlin, 4.9.2014

Lena Bonsiepen
Wolfgang Coy

Exzellenzcluster „Bild-Wissen-Gestaltung“
der Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

http://www.bwg.hu-berlin.de
lena.bonsiepen@hu-berlin.de
coy@hu-berlin.de

B.2

Humanist-Computer Interaction Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik

Manuel Burghardt, Christian Wolff – Institut für Information und Medien, Sprache und Kultur, Fakultät für Sprach-, Literatur- und Kulturwissenschaften, Universität Regensburg

Humanist-Computer Interaction

Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik

Autoren

Manuel Burghardt & Christian Wolff
Lehrstuhl für Medieninformatik
Institut für Information und Medien, Sprache und Kultur (I:IMSK)
Fakultät für Sprach-, Literatur- und Kulturwissenschaften
Universität Regensburg, 93040 Regensburg
{manuel.burghardt, christian.wolff}@ur.de

Ausgangsargument

Wir gehen von folgendem Ausgangsargument aus: Arbeitstechniken in den Digital Humanities sind geprägt von der Verfügbarkeit und der Nutzung interaktiver Softwarewerkzeuge („Tools"). Aspekte der *Usability* (Gebrauchstauglichkeit eines Produkts) und der *User Experience* (gesamtes Nutzererlebnis) spielen eine nicht unwesentliche Rolle für die erfolgreiche und effiziente Einführung und Etablierung digitaler Arbeitstechniken. Die dazu erforderlichen Entwicklungsmethoden sind bisher allerdings kaum etabliert, ähnlich wie man dies allgemeiner auch für die Methoden des Software Engineering in den Digital Humanities sagen kann. Hinzu kommt, dass viele Arbeitsprozesse in den Digital Humanities zwar werkzeug*gestützt* ablaufen, aber der intellektuelle Input des Forschers weiterhin von zentraler Bedeutung ist, und gerade bei großen Datenumfängen viele Ressourcen in Anspruch nimmt (z. B. bei der Annotation linguistischer Daten; bei der Erfassung qualitativer Daten; bei der Klassifikation von Beobachtungen; etc.). Dabei können besser benutzbare Werkzeuge zu nicht unerheblichen Skaleneffekten führen, was letztlich auch ein ökonomisches Argument für besser bedienbare Tools liefert.

Vielfalt der Werkzeuge und deren Defizite im Bereich Usability

Mittlerweile existiert eine schon kaum mehr überschaubare Menge an Tools[1], die belegen, dass digitale Arbeitspraktiken längst in der Welt der Geisteswissenschaften angekommen sind. Dabei fällt auf, dass sehr viele der einschlägigen Tools nur bedingt oder gar nicht auf Aspekte des *Interface Design* und des *Usability Engineering* eingehen. Offenkundig sind viele der Applikationen ohne tech-

[1] Vgl. etwa das *Digital Research Tools* (DiRT) Wiki für eine umfangreiche Sammlung digitaler Tools für Geisteswissenschaftler: http://dirtdirectory.org/ (letzter Zugriff am 10.8.2014).

nisches Vorwissen und zumindest grundlegende Kenntnisse in Feldern wie Markup-Sprachen, Datenbanken und Programmierung kaum zu bedienen. Ein bekanntes Beispiel ist etwa das *Python NLTK*[2], ein umfangreiches Softwarepaket zur Verarbeitung von Sprachdaten, welches aber nur aus der Kommandozeile bedient werden kann, und damit für viele Sprachwissenschaftler eine erhebliche Hürde darstellt. Aus Sicht der Medieninformatik, die sich intensiv mit den Themen HCI (*Human-Computer Interaction*) und UX (*User Experience*) beschäftigt, ist somit ein wesentliches Desideratum an die Digital Humanities, die Berücksichtigung bestehender Usability Engineering-Praktiken. Dadurch wird die Einstiegshürde für technisch weniger versierte Nutzer gesenkt, und die Attraktivität digitaler Tools für die Zielgruppe der Geisteswissenschaftler erhöht.

Zielgruppenorientierte Entwicklung als besonderes Problem

Gleichzeitig scheinen viele bestehende Tools an der tatsächlichen Zielgruppe „vorbeientwickelt" zu werden (Warwick et al., 2008; Warwick 2012; Gibbs & Owens, 2012): Digital Humanities-Tools müssen demnach nicht nur allgemeine *Best Practices* aus dem Bereich der Usability-Forschung umsetzen, sondern auch auf besondere Charakteristika geisteswissenschaftlicher Anwender eingehen, indem Arbeitstechniken und -praktiken aus dem nicht-digitalen Arbeitsalltag möglichst gut im digitalen Interface abgebildet werden. Dazu gehören zum einen die Berücksichtigung von etablierten Metaphern und Workflows, und zum anderen die Erhebung konkreter Anforderungen und Erwartungen der Nutzer an ein entsprechendes Tool.

Konkrete Beispiele aus eigener Arbeit

Am Lehrstuhl für Medieninformatik an der Universität Regensburg finden sich zahlreiche Forschungsprojekte[3], die sich der Frage widmen, wie digitale Tools durch angemessene Interfaces für Geisteswissenschaftler attraktiv und letztlich praktisch nutzbar gemacht werden können. Beispielhaft zu nennen sind hier:

- *WebNLP* (Burghardt et al., 2014): Web-Interface für die Verwendung von *Python NLTK*-Funktionen und Visualisierung mithilfe von *Voyant Tools*
- *TWORPUS* (Bazo et al., 2013): Tool mit grafischer Benutzeroberfläche zur Erstellung von Tweet-Korpora
- *To See or Not to See* (Wilhelm et al., 2013): Interaktives Web-Interface zur Visualisierung quantitativer Aspekte von Shakespeare-Dramen

[2] Verfügbar unter http://www.nltk.org/ (letzter Zugriff am 10.8.2014).
[3] Ein Überblick zu aktuellen Projekten findet sich auf dem Forschungsblog *Digital Humanities Regensburg*: http://dhregensburg.wordpress.com/dh-projekte/ (letzter Zugriff am 31.8.2014).

Pragmatische Sicht auf das Verhältnis von Informatik und Digital Humanities

Das Verhältnis zwischen Geisteswissenschaften und digitalen Methoden aus der Informatik war und ist Gegenstand intensiver Diskussionen (vgl. etwa Ramsay, 2011; Gold, 2012; Berry, 2012), was sich zuletzt auch bei der DHd 2014 in Passau gezeigt hat. Wir plädieren für eine pragmatische Sicht auf das Feld der Digital Humanities, in welcher die Existenz einer Vielzahl bestehender DH-Tools bereits als gegeben angesehen werden kann, und für die weitergehende Aspekte der *guten* Software-Entwicklung in den Vordergrund gerückt werden sollten. Pragmatisch auch insofern, als wir nicht davon ausgehen, dass es genau einen idealen Weg für die Synthese informatischer und geisteswissenschaftlicher Methoden gibt.

Plädoyer für das Forschungsfeld *Humanist-Computer Interaction*

Während Gerhard Heyer in seinem Vortrag auf der DHd 2014 bereits das Aufgreifen bestehender Software Engineering-Praktiken in den Digital Humanities thematisierte, so schlagen wir zur Steigerung der Akzeptanz solcher Tools die Berücksichtigung von Erkenntnissen aus dem Bereich des Usability Engineering als besonderen Schwerpunkt vor. Soweit man Usability Engineering als Teilgebiet des Software Engineering versteht, ist dies ein Teilaspekt der allgemeineren Forderung nach besseren Software-Entwicklungspraktiken in den Digital Humanities. Unter Berücksichtigung der besonderen Nutzergruppe der Geisteswissenschaftler können diese neuen Anforderungen an die Digital Humanities auch unter dem Begriff der *Humanist-Computer Interaction* zusammengefasst werden.

Ausblick

Aus dem hier skizzierten Feld der *Humanist-Computer Interaction* ergeben sich weiterhin eine ganze Reihe interessanter Forschungsfragen, nur einige seien als Ausblick genannt:

- Wie kann eine Bestandsaufnahme zum Stand der UX-Qualität von DH-Tools durchgeführt werden?
- Gibt besonders gute / schlechte Tools?
- Gibt es „leichtgewichtige“ / „Discount“- / „Guerilla“-Methoden für die Einführung von UX-Methoden in den Digital Humanities, die besonders schnellen Erfolg versprechen?
- Lässt sich der ökonomische Nutzen, bessere UX-Qualität im DH-Bereich, quantifizieren / belegen? Lassen sich Drittmittelgeber dadurch beeindrukken?

Quellen

Bazo, Alexander Burghardt, M., & Wolff, C. (2013). TWORPUS – An Easy-to-Use Tool for the Creation of Tailored Twitter Corpora. In I. Gurevych, C. Biemann, & T. Zesch (Eds.), *Proceedings of the 25th International Conference of the German Society for Computational Linguistics and Language Technology, GSCL '13* (pp. 23–34). Heidelberg: Springer.

Berry, D. M. (2012). *Understanding Digital Humanities*. Palgrave Macmillan.

Burghardt, M., Pörsch, J., Tirlea, B., & Wolff, C. (2014). WebNLP: An Integrated Web-Interface for Python NLTK and Voyant. In *Proceedings of the 12th KONVENS 2014* (to appear).

Gibbs, F., & Owens, T. (2012). Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly (Online Journal), 6*(2).

Gold, M. K. (2012). *Debates in the Digital Humanities*. Univ Of Minnesota Press.

Ramsay, S. (2011). *Reading Machines: Toward an Algorithmic Criticism*. Urbana, Chicago. and Springfield: University of Illinois Press.

Warwick, C. (2012). Studying users in digital humanities. In C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 1–21). London: Facet Publishing.

Warwick, C., Terras, M., Huntington, P., & Pappa, N. (2008). If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Literary and Linguistic Computing*, *23*(1), 85–102.

Wilhelm, T., Burghardt, M., & Wolff, C. (2013). "To See or Not to See" - An Interactive Tool for the Visualization and Analysis of Shakespeare Plays. In R. Franken-Wendelstorf, E. Lindinger, & J. Sieck (Eds.), *Kultur und Informatik: Visual Worlds & Interactive Spaces* (pp. 175–185). Glückstadt: Verlag Werner Hülsbusch.

B.3

How to integrate conservation data in the Linked Open Data Cloud

Ernesto William De Luca – Fachhochschule Potsdam

How to integrate conservation data in the Linked Open Data Cloud.

Ernesto William De Luca (deluca@fh-potsdam.de)

In conservation the documentation is an integral part of the restorers' work to describe their restored objects. It is important and indispensable to understand the phases of conservation and to allow decision-making for conservation and conservation procedures. The information they contain about the conducted conservation procedures, as well as the methods and the materials used are an immeasurable store of knowledge for this discipline and build the basis of the knowledge used for the conservation of a given object.

Documentation is very time consuming; conservators have to categorize photos, give detailed description of the procedures and justify decisions about materials. Every conservator delivers his/her own documentation to the respective monument administrations that archive it in their shelves, without any possibility to access them digitally. In this case, the access to already conducted procedures, as well as the provision of the knowledge about the methods and materials used are very difficult, often almost impossible. At the moment, it is not possible to link different documentations that contain information about similar objects, methods or materials used in the same way. These properties are not recognizable, because of this missing link between them.

Ontologies and Cultural Heritage. Ontologies are the central components of semantic tools and have been present in the art and cultural area for some time. For instance, the CIDOC Conceptual Reference Model (CRM) supports the integration, communication and the exchange of differently structured information from the cultural heritage field. Meanwhile, there are some projects that already use these semantic technologies. A project of the British Museum (http://collection.britishmuseum.org), allows the collection of all available data as Linked Open Data (LOD). Also, the currently implemented EU project Europeana (http://www.europeana.eu/portal/) focuses on the global accessibility of cultural information as LOD. According to the EU Commission, a prominent role for the future of the European knowledge and information society is played by sharing the electronic content. In 2009, the German Federal Government had decided to create a German Digital Library (Deutsche Digitale Bibliothek, www.deutsche-digitale-bibliothek.de), which should link numerous German cultural and scientific facilities and integrate them at European level in the Europeana.

The field of conservation is, in this context, still underrepresented. The only semantic contribution in this domain is the Australian project "The Twentieth Century in Paint" (http://www.20thcpaint.org/), which investigates paint materials used in art works of the 20th Century in the Asian-Pacific region. One goal of the project is to develop a suitable ontology. This was developed for a specific sub- domain and can be reused, but represents only a very small area of conservation.

To develop a generic conservation ontology, we have to consider the numerous measures and materials, as well as the damage patterns and causes of loss, which can vary greatly depending on the materiality and exposure of the object to be restored. To achieve a high acceptance in the professional world, conservation experts must be involved in this development process. Ontologies in the field of conservation documentation have not yet been developed. This is a big gap in the research infrastructure of conservation and other fields of cultural heritage.

In cooperation with the Cologne Institute of Conservation Sciences (specifically with Prof. Dr. Peter Kozub), we developed a system (http://semres.fh-potsdam.de) that bases on an own ontology for conservation documentation that includes different semantic relations. The ontology is developed as a basis for semantic processing of documentation for conservators. The TripleStore is based on Erlangen CIDOC Conceptual Reference Model (CRM, http://erlangen-crm.org/current-version) and has been extended with VIAF (http://viaf.org/) and Geonames (http://www.geonames.org).

The resulting research is a joint effort of conservators, information and computer scientists in order to help conservators in organizing and retrieving their own documentation semantically. It aims to build on European initiatives and efforts that seek to free access of all cultural heritage information. The prior objective of this work is the preparation of data for the restorers that are automatically linked in the Linked Open Data Cloud. In this way, the access and reuse of scientific data on restoration is made possible not only for conservators, but also for architects, historians, archeologist and interested people.

In the final paper, we will present the system architecture and describe in more details how the ontology has been built. We will give some more concrete examples, in order to make clear how the system uses the information stored in the relational database and in the triple store. The described resources will be made available, as well as the classification of 539 material classes (related mostly to stone conservation).

B.4

Stemmatology, an interdisciplinary endeavour

Armin Hoenen – Goethe-Universität Frankfurt am Main

Stemmatology, an interdisciplinary endeavour

Armin Hoenen

31-08-2014

Abstract

In this paper, stemmatology for the digitisations of ancient documents is focussed as an example of a digital humanistic approach. A *stemma codicum* is a visualisation of the genealogical relationships between manuscripts containing a historical text, which was typically transmitted chirugraphically. The visualisation is arising from the copying history of these texts and is ususally an acyclic directed graph with one root node. While the humanities have developed various theories and practises closely connected to the goal of producing a textual edition (see for instance West (1973)), for computer science the field remains relatively recent and is dominated by the humanities. Computationally produced stemmata are revised according to philological expertise (see for instance Roelli and Bachmann (2010)) and few new editions use computationally generated stemmata. Available encoded metadata (e.g. paleographic) is still largely missing. In fact, the humanities have a vast research advantage in terms of research time: As was declared as early as 1734 by a scholar named Bengel: "a perfect edition of the New Testament would propose a classification of the codices for their genealogical relations", Pasquali (1988). Since then, most editions start to give a stemma in their preface or appendix. Before this, according to Timpanaro (2005), tables had been in use. There is no explicit standard yet for the visual representation of stemmata, but some general practises such as denoting lost manuscripts at internodes (reconstructed hypearchetypes) by Greek letters and minor influences across manuscripts by dotted lines are visible.

While West (1973) remarks that "the time has not yet come when manuscripts can be collated automatically" and "it seems, computers can serve us best by making concordances and the more unsub-

tle kinds of metrical analysis", in the later half of the 1980ies automated stemmatology loaned methods, algorithms and software from bio-informatic phylogeny and started to be used in textual criticism, Christopher J. Howe and Windram (2012). To date, the methods are not fully adapted for the needs of textual criticism. As an example, a large number of representations are exclusively bifurcating structures because new species usually do not split off simultaneously (apart from some cases in adaptive radiation) while manuscripts are usually not copied just once. The adaptation of philological needs into stemmatology is an ongoing digital humanities process.

Mehler and Lücking (2014) discuss the gap between Computation and Humanities. In their paper, they mention the stemmatological reconstruction of lost hypearchetypes as an exemplary manifestation of a more general problem found among others in text classification, topic detection and labeling, textual entailment, topic tracking and text reuse. Accordingly, stemmatic problems are believed here to be generalizable and a valid primer for similar problems in bio-informatics and other fields. For instance, bacterial lateral gene transfer is a potentially similar problem to the contamination problem (one manuscript having two *vorlages*). Summarizing, a subfield which formerly had two participating sciences, biology and computer science, in this case now gains a third partner.

Generally speaking, the answer to the question who transfers the models is multiple, there is no rule as to whether this must be a computer scientist or an adherent of the humanities discipline, a "half-blood" or a team of two or more scientists. Ideally, the digital humanities approach as an interdisciplinary approach can integrate multiple advantages. The computer scientific advantages of effective data processing and the results from statistical points of view onto the data can combine with the knowledge for detail and historical context of the humanists. As a broader example, Rayner et al. (2012) rejected letter ngram based reading modelling as neurologically plausible simulations on the basis of psycholinguistic research, even though they had some predictive power. This points to a possible digital humanities compromise, which is to take that solution which gives the best results (computation) but is at the same time plausible (humanities).

The full paper will contain a comprehensive history of the development of stemmatology and abstracting from that it will highlight the co-involved digital humanities dimension.

References

Christopher J. Howe, R. C. and Windram, H. F. (2012). Responding to criticism of phylogenetic methods in stemmatology. *SEL studies in English Literature*, 52(DOI: 10.1353/sel.2012.0008):51–67.

Mehler, A. and Lücking, A. (2014). On Covering the Gap between Computation and Humanities. *Dagstuhl Reports*. forthcoming.

Pasquali, G. (1988). *Storia della tradizione e critica del testo*. Casa editrice Le lettere, Firenze.

Rayner, K., Pollatsek, A., Ashby, J., and jr. Clifton, C. (2012). *Psychology of Reading*. Psychology Press, New York/Hove.

Roelli, P. and Bachmann, D. (2010). Towards generating a stemma of complicated manuscript traditions: Petrus alfonsi's dialogus. *Revue d'histoire des textes*, 5(4):307–321.

Timpanaro, S. (2005). *The Genesis of Lachmann's Method*. University of Chicago, Chicago.

West, M. L. (1973). *Textual Criticism and Editorial Technique: Applicable to Greek and Latin texts*. Teubner, Stuttgart.

B.5

Zum Stand der Interdisziplinarität in den Digital Humanities: Perspektiven aus dem Bereich der Social-Media-Forschung

Katharina Kinder-Kurlanda, Katrin Weller –
GESIS Leibniz-Institut für Sozialwissenschaften

Zum Stand der Interdisziplinarität in den Digital Humanities: Perspektiven aus dem Bereich der Social-Media-Forschung

Katharina Kinder-Kurlanda, Katrin Weller

In diesem Beitrag betrachten wir die Digtal Humanities (DH) als interdisziplinäres Projekt und möchten die im Call aufgeworfene Frage danach, wie Geisteswissenschaften und Informatik im Rahmen dieses Forschungsprogrammes produktiv zusammenarbeiten können, adressieren. Unter welchen Umständen eine Zusammenarbeit besonders vielversprechend scheint, stellen wir am Beispiel der Social-Media-Forschung dar. Social-Media-Dienste und ihre Nutzer rücken zurzeit in verschiedenen disziplinären Kontexten in den Blickpunkt der Forschungsaktivitäten. Beispiele umfassen unter anderem die Untersuchung von Wahlkampf-Kommunikation in der Politikwissenschaft, Auswertungen zum Online-Sprachgebrauch in der Linguistik oder die Anwendung von Netzwerk-Modellen in Informatik und Physik. In einem laufenden Projekt erarbeiten wir derzeit eine Bestandsaufnahme zu Perspektiven und Methoden der Social-Media-Forschung. Hierzu führen wir qualitative Experteninterviews mit Wissenschaftlern durch, die verschiedene Forschungsschwerpunkte verfolgen, aber jeweils Daten aus Social-Media-Diensten für ihre Forschung nutzen.

In diesem Zusammenhang spielen Erfahrungen mit Interdisziplinarität eine große Rolle. Vielfach trafen wir auf konkrete Projekte interdisziplinärer Zusammenarbeit, die versuchen, informatische Verfahren der Berechenbarkeit mit geistes- sozial- und medienwissenschaftlichen Ansätzen zu verknüpfen. Dies liegt zu einem großen Maß in den Eigenschaften des Forschungsgegenstands „Social Media“ begründet. Aus geistes- und sozialwissenschaftlicher Perspektive stellt sich die Arbeit mit Social-Media-Daten oftmals als „Big Data“ dar, die nicht ausschließlich mit traditionellen Verfahren erarbeitet werden sollte: die Zusammenarbeit mit anderen Fachbereichen verspricht unterschiedliche Blickpunkte und Herangehensweisen zur Auseinandersetzung mit dem Thema und damit einen zusätzlichen Erkenntnisgewinn. Die Social-Media-Forschung eignet sich daher hervorragend für eine Betrachtung der Rolle von Interdisziplinarität in der Erschließung der Potentiale von Big Data, wobei unter anderem sozial- und geisteswissenschaftliche sowie mathematisch-informatische Herangehensweisen zu betrachten sind.

Im Folgenden stellen wir zentrale Ergebnisse unseres Projektes vor. In einer ersten Projektphase haben wir uns speziell mit geistes-, sozial-, und medienwissenschaftlich orientierten Social-Media-Forschenden beschäftigt. Dies ermöglicht es uns nun, auszuwerten, wie diese ihre Erlebnisse in der interdisziplinären Zusammenarbeit insbesondere mit Informatikern schilderten und worin aus ihrer Sicht Potentiale aber auch Schwierigkeiten beim Einbeziehen der spezifischen Ansätze der Informatik in den jeweiligen Projekten lagen[1]. Ziel des vorliegenden Beitrages ist es, die daraus gewonnenen Erkenntnisse in die Diskussion um das Potential der Informatik in den DH einzubringen und diese aus Sicht der von uns befragten Forschenden zu beleuchten. Dabei handelt es sich sicherlich um eine spezielle Fallstudie, da die Beobachtung zumeist sozialer Prozesse (‚Social‘ Media) oft eher den Sozial- als den Geisteswissenschaften zugerechnet wird. Es geht uns jedoch hier darum, Erkenntnisse einzubringen, die wir für den Bereich der weiteren DH übertragbar halten.

Potentiale und Schwierigkeiten interdisziplinärer Projektzusammenarbeit

Die von uns befragten Social-Media-Forschenden hatten zumeist bereits damit begonnen, sich Programmier- oder andere technische Kenntnisse zu eigen zu machen um mit Social-Media-Daten arbeiten zu können, meist erwarben sie diese Kenntnisse durch eigene Initiative; alle Befragten stimmten darin überein, dass **Social-Media-Forschung informatische Kenntnisse benötigt**.

[1] Die hier dargestellten Ergebnisse basieren auf zwanzig qualitativen Interviews mit Social Media Forschenden aus Europa, Amerika und Australien. Eine umfassendere Darstellung des Forschungsprojektes ‚The Hidden Data of Social Media Research‘ und des Themas Interdisziplinarität findet sich in: Kinder-Kurlanda & Weller (2014).

Von uns befragte Social-Media-Forschende drückten zudem die Hoffnung aus, dass die Informatik dabei helfen könne, **vormals unlösbare Probleme, speziell in der Big Data Forschung, zu bearbeiten**. Zum Beispiel können herkömmliche medien- oder sozialwissenschaftliche Ansätze Phänomene der Meme-Bildung nicht abbilden, da "questions of spread are really hard to get to from a qualitative point of view and even from a quantitative standpoint. I think that it would take some really big systematic work" (so eine der Befragten). Herangehensweisen, die - jenseits des Einsatzes von Skripten und plattformspezifischen APIs - komplexere Programmierkenntnisse, höhere Rechenleistung oder neue Ansätze in der Informatik beinhalten, schienen dann besonders hilfreich, wenn Probleme mit hoher Komplexität, wie sie z.B. durch eine Betrachtung verschiedener Social-Media-Plattformen und Formate entsteht, vorlagen.

Im Hinblick auf die interdisziplinäre Zusammenarbeit konnten wir feststellen, **dass Kollaborationen unmöglich werden können, wenn Vorstellungen über Methodologie und Epistemologie zu sehr voneinander abweichen** oder nicht kompatibel sind. Ein Beispiel hierfür ist eine bestimmte Art der qualitativen Datenerhebung, wie sie auch in den Geisteswissenschaften sehr verbreitet ist: wird das Erheben der Daten bereits als Teil des Erkenntnisprozesses gesehen, schließt dies per Definition aus, dass auch Dritte die Analyse dieser Daten vornehmen können. Epistemologische Differenzen sind ein für die Zusammenarbeit zwischen Geistes- oder Sozialwissenschaftlern und Informatikern typisches Problem, da bestimmte Forschungs-Paradigmen nur schwer miteinander vereinbar sind. Wenn beispielsweise zahlreiche Projekte in der Informatik darauf aufbauen, Systeme oder Lösungen auf Basis einer neuen Annahme über die Welt aufzusetzen, so widerspricht dies hermeneutischen wie auch konstruktivistischen Forschungsansätzen, nach denen Begriffsdefinitionen und die Beobachtung und Infragestellung neuer Annahmen selbst Ziel der Forschung ist[2]. Gerade das Ziel der DH, technische Methoden und hermeneutische Paradigmen zu vereinen scheint auch an ein praktisches Hindernis in Bezug auf die in den jeweiligen Disziplinen publizierbaren Ergebnisse zu stoßen, die jedoch oft innerhalb von Projekten die strategischen Entscheidungen über das Vorgehen bestimmen.

Oft stellte sich als Schwierigkeit innerhalb der Projektzusammenarbeit zudem die Tatsache heraus, dass nicht auf Augenhöhe zusammengearbeitet wurde, beispielsweise wenn Informatikern im Social-Media-Projekt ausschließlich die aus Sicht informatischer Forschungsinteressen wenig attraktive Aufgabe des ‚Datensammelns' zugeteilt wurde[3]. Umgekehrt beschreiben die eher sozial-und geisteswissenschaftlichen Social-Media-Forschenden wiederum eine Art **Problemorientiertheit der Informatik**-Kollegen: Diese sähen Nicht-Informatiker im Projekt häufig als "an end user that they need to provide a technological solution for", und nicht als gleichberechtigte Partner in einem gemeinsamen Forschungsprojekt mit dem Ziel des Weiterentwickelns und Infragestellens von Methoden und Paradigmen.

Erwartete Beiträge für die Social-Media-Forschung aus der Informatik

Besonders attraktiv für Social-Media-Forschende in unserer Befragung war die **Verfügbarkeit 'neuer' Daten** ("Oh my gosh, we have this amazing data!"). Mithilfe von informatischen Methoden gesammelte (gecrawlte) und bearbeitete Daten sind einfacher und schneller zu erlangen als etwa Umfrage- oder Interviewdaten und haben zudem den Vorteil, dass Forschungsfragen angepasst und während der Erhebung erweitert oder verändert werden können. Die oft aus solchen Ansätzen folgende ‚datengetriebene' Forschung und das Problem eines naiven Induktivismus ist allerdings auch einer der größten Kritikpunkte an der sich entwickelnden Big Data / Social-Media-Forschung – sowohl bei den von uns befragten Social-Media-Forschenden selbst als auch in der Literatur (siehe z.B. Frické 2014).

Die **Möglichkeit zur unmittelbaren Beobachtung von in situ - Kommunikation und Ereignissen,** die durch das Beforschen von Social-Media-Daten möglich wird, wurde ebenfalls geschätzt. Vormals nur schwer oder unmöglich zu beobachtende Aspekte menschlicher Aktivität werden analysierbar - etwa

[2] Vgl. zu diesem Thema ausführlicher: Kinder-Kurlanda (2014)
[3] Heyer et al. (2014) beschreiben dieses Problem ebenfalls.

Diskussionen über politische Ereignisse oder Proteste. Damit werden durch die Verfügbarkeit von Social-Media-Daten **neue Möglichkeiten zur Beobachtung gesellschaftlicher Prozesse eröffnet**. Das Verhältnis zwischen Internet und Gesellschaft wird von den Forschenden dabei als komplex wahrgenommen: das Internet ermöglicht die Beobachtung der Gesellschaft, diese beobachtet sich jedoch auch selbst „for the first time (…) in a new way" und wird wiederum von dieser Selbstbeobachtung beeinflusst.

Interdisziplinarität als Herausforderung in den DH

Die interdisziplinäre Zusammenarbeit im Rahmen der Social-Media-Forschung stellte sich in unserer Untersuchung dann als besonders schwierig heraus, wenn unterschiedliche Ideen über Validität oder Aussagekraft von Datenkorpora, über best practice, oder über Forschungsinteressen bestanden. Unterschiedlichen Standpunkte stellen ein typisches Hindernis in der interdisziplinären Zusammenarbeit dar (Brewer 1999), scheinen jedoch gerade in der Bewertung sowohl der neu verfügbaren Daten als auch der neu angewendeten (informatischen) Methoden oft zu kritischen Streitpunkten in interdisziplinären Kollaborationen zu werden.

Wir folgern aufgrund der Ergebnisse unserer Forschung dass, um ein erfolgreiches Zusammenarbeiten von Informatik und Geisteswissenschaften im Rahmen der DH zu ermöglichen, epistemologische Grundlagen diskutiert werden müssen. Innerhalb von Projekten ist eine erfolgreiche interdisziplinäre Zusammenarbeit nur dann möglich, wenn die Art und Weise wie innerhalb der angewendeten Methoden eine in den jeweiligen Disziplinen akzeptierte Validität der Forschung erreicht wird, übereinstimmen – oder sich zumindest nicht widersprechen.

Literatur:

Brewer, G.D. (1999): The challenges of interdisciplinarity. Policy Sciences 32: 327-337.

Frické, M. (2014): Big data and its epistemology. Journal of the Association for Information Science and Technology JASIST, Article first published online: 2 May 2014, DOI: 10.1002/asi.23212

Heyer, G., Niekler, A., Wiedemann, G. (2014): Brauchen die Digital Humanities eine eigene Methodologie? Überlegungen zur systematischen Nutzung von Text Mining Verfahren in einem politikwissenschaftlichen Projekt. In: 1. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHd 2014).

Kinder-Kurlanda, K. E., Weller, K. (2014): "I always feel it must be great to be a hacker!": The role of interdisciplinary work in social media research. In *Proceedings of the 2014 ACM conference on Web Science*, 91-98. New York: ACM. Online: http://www.sheridanprinting.com/14-websci4chRV610jmp/docs/p91.pdf.

Kinder-Kurlanda, K. E. (2014): Ethnography in a computer science centered project. ICC'14 Workshop, ACM WebSci'14 Conference, Bloomington. Online: http://www.icc.ecs.soton.ac.uk/papers/icc14_submission_2.pdf.

Über die Autorinnen:

Katharina Kinder-Kurlanda ist Sozialwissenschaftlerin und arbeitet seit mehr als zehn Jahren im Grenzgebiet zwischen Kulturanthropologie und Informatik. Sie forscht interdisziplinär zur Epistemologie von Big Data, zu Empirical Secure Software Engineering und zum Internet of Things und leitet das Secure Data Center beim GESIS Leibniz Institut für Sozialwissenschaften. Ihren PhD erhielt sie von der Lancaster University , wo sie eine Ethnografie im Rahmen eines interdisziplinären Projektes zu Ubiquitous Computing durchführte. Sie hat in Tübingen und Frankfurt am Main Kulturanthropologie, Informatik und Geschichte studiert.

Katrin Weller ist Informationswissenschaftlerin mit Forschungsschwerpunkten im Bereich Web Science, Social-Media-Analyse und Science 2.0 und beim GESIS Leibniz-Institut für Sozialwissenschaften für Social-Media als sozialwissenschaftliche Forschungsdatenquelle zuständig. Sie studierte Geschichte, Medienwissenschaft und Informationswissenschaft an der Heinrich-Heine-Universität Düsseldorf und promovierte dort 2010 zum Thema „Knowledge Representation in the Social Semantic Web". Sie verfügt über langjährige Erfahrung in der Zusammenarbeit mit ForscherInnen verschiedener Disziplinen wie (Bio-)Informatik, Linguistik, Erziehungswissenschaft und Jura.

B.6

Abstract zum Workshop „Informatik für die Digital Humanities“ Eine Betrachtung aus dem Blickwinkel der Visualisierungsforschung

Steffen Koch – Institut für Visualisierung und Interaktive Systeme, Universität Stuttgart

Abstract zum Workshop „Informatik für die Digital Humanities“
Eine Betrachtung aus dem Blickwinkel der Visualisierungsforschung

Autor: Dr. Steffen Koch, Institut für Visualisierung und Interaktive Systeme, Universität Stuttgart

Visualisierungsforschung im Kontext dieses Abstracts sehe ich als eine anwendungsorientierte Informatikdisziplin. Sie ist per se interdisziplinär und zielt darauf ab, Daten die in anderen wissenschaftlichen Disziplinen bzw. Anwendungsgebieten anfallen, aufzubereiten und in geeigneter Form, meist aggregiert, darzustellen. Dabei dient Visualisierung nicht mehr ausschließlich der Präsentation von Ergebnissen, sondern wird zunehmend ein interaktives Instrument zur Exploration und zur Steuerung von zielgerichteten Analysen durch die Benutzer.

Die Geistes- und Sozialwissenschaften haben den Nutzen von Visualisierung bereits seit einiger Zeit für sich entdeckt, wobei Visualisierung hier sicherlich näher in ihrem ursprünglichen Wortsinn verstanden und damit deutlich breiter definiert wird, als das in der gleichnamigen Informatikdisziplin mittlerweile der Fall ist. Während unter letzterer primär eine datengetriebene, algorithmisch generierte Form von Visualisierung verstanden wird, schließt diese allgemeinere Betrachtung auch statische, manuell erzeugte Formen mit ein.

Aus den Erfahrungen der letzten Jahre, den bereits mit Visualisierungsunterstützung erzielten Erfolgen in den Digital Humanities, sowie dem weiterhin ungebrochenen Interesse, das der Visualisierungsforschung von Seiten der Geistes- und Sozialwissenschaften zu Teil wird, steht aus meiner Sicht die zukünftige Fortführung solcher Zusammenarbeiten nicht in Frage. Dies schließt insbesondere die Nutzung von Visualisierung als Teil hermeneutischer Betrachtungen mit ein, denn Visualisierung ist bereits jetzt ein wichtiges Bindeglied zwischen automatischer Datenaufbereitung und der menschlichen Wahrnehmung und damit dem Verstehen und der Interpretation solcher automatischer Verfahren. Allerdings erfordert auch der richtige und sinnvolle Einsatz von Visualisierung einen gewissen Kenntnisstand des Interpretierenden oder eine Art von „visueller Belesenheit“, wenn man so möchte. Sonst kann es leicht passieren, dass z.B. visuellen Artefakten, die einer automatischen Bearbeitung entstammen, Bedeutungen beigemessen werden, die gar nicht vorhanden sein können.

Dennoch gestaltet sich die Zusammenarbeit in Projekten, an denen Geisteswissenschaftler und Informatiker beteiligt sind, oft schwieriger als dies in anderen interdisziplinären Projekten der Fall ist. Mit Sicherheit gibt es zunächst eine Kommunikationsbarriere, die es generell in interdisziplinären Projekten gemeinsam zu überwinden gilt. Das ist in Kooperationen zwischen beispielsweise Naturwissenschaftlern und Informatikern nicht sehr anders. Und doch scheinen die Barrieren höher zu sein, was sicherlich nicht zuletzt in den „wissenschaftskulturellen“ Unterschieden liegt.

Ein aus meiner Sicht großes Problem stellt der Mangel an informatischem Grundinstrumentarium und –wissen seitens der Geisteswissenschaften dar. Das ist natürlich eine starke Verallgemeinerung und für ganze Fachgebiete, die schon lange mit Rechnerunterstützung forschen, wie z.B. die Sozial- und Politikwissenschaften, unzutreffend. Dennoch scheint es in vielen anderen geisteswissenschaftlichen Disziplinen diese Grundausstattung nicht zu geben. Selbstverständlich gilt das auch umgekehrt für das Vorhandensein von geisteswissenschaftlichem Grundwissen und dem Verständnis von geisteswissenschaftlichen Forschungsfragestellungen. Ablehnung, wie C.P. Snow sie

beschrieben hat, gibt es sicherlich heutzutage auch, aber insgesamt scheint mir doch das Interesse und der ernstgemeinte Wille zur Kooperation in Geisteswissenschaft und Informatik zu überwiegen.

Vergleicht man interdisziplinäre Kooperation, wie z.B. zwischen Naturwissenschaftlern und Visualisierern, so stellt sich üblicherweise die Situation so dar, dass eine Visualisierungslösung gesucht wird, die über das hinausgeht, was die jeweilige Fachdisziplin mit ihren vorhandenen Standardvisualisierungswerkzeugen erreichen kann. Daraus ergibt sich geradezu ein Projektziel, das über den Stand der Technik hinausgeht und damit schon fast automatisch einen forschungsrelevanten Fortschritt beinhaltet, der auch publizierbar und damit, wie heutzutage in vielen Informatikfächern üblich, als anerkannte Forschungsleistung bewertbar ist.

Dies ist in Digtital Humanities Projekten nach meiner Erfahrung nicht immer einfach möglich. Hier besteht häufig die Situation, dass aus Sicht der Geisteswissenschaften zunächst Probleme gelöst werden müssen, die aus Informatikperspektive keinen oder nur sehr geringen Forschungsbezug haben. Es steht außer Frage, dass durch die Bereitstellung oder Anpassung vorhandener oder naheliegender Lösungen ein DH-Fortschritt erzielt wird und Ergebnisse in dieser Community publiziert werden können, aber die Frage ist, ob beispielsweise ein Umsetzer aus der Informatik damit einen überzeugenden Beitrag zu seiner Informatik-Dissertation leisten kann.

Andererseits beschäftigen sich die Geisteswissenschaften, wie im Call-For-Papers für diesen Workshop treffend bemerkt, oft mit schwer zu formalisierenden und mit den gegenwärtigen in der Informatik entwickelbaren Ansätzen möglicherweise schwer oder gar nicht lösbaren Problemen. Berücksichtigt man all diese Aspekte, dann ergibt sich zwar ein enormer Bedarf an Informationstechnologischen Lösungen in den Digital Humanities, aber nur ein relativ schmaler Bereich, in dem eine Win-win-Situation hinsichtlich der Forschung möglich ist, ohne dass entweder die geisteswissenschaftliche Fragestellung zu stark simplifiziert werden muss oder der Informatikbeitrag nicht über den aktuellen Stand der Technik hinaus geht.

Mein Austausch mit Informatikern aus den Bereichen der maschinellen Sprachverarbeitung, des Data Mining, der Datenbankforschung und anderen Informatikdisziplinen, sowie mit Geistes- und Sozialwissenschaftlern, die in die entsprechenden Digital Humanities Projekte involviert sind oder an zukünftigen Kooperationen Interesse zeigen, legt nahe, dass die zuvor geschilderten Beobachtungen kein Einzelfall sind.

Aus meiner Perspektive spricht daher Vieles dafür, dass tatsächlich neue dedizierte Studiengänge mit Fokus auf Digital Humanities geschaffen werden müssen, die eine Ausbildung in beiden „Kulturen“ mit wählbaren Schwerpunkten ermöglichen. Dies würde mittelfristig nicht nur zu einer höheren Verfügbarkeit von Software zur Lösung von gewissen Standardproblemen on den DH führen, sondern auch eine Anpassung der als Forschungsleistung anerkannten Publikationen/Publikationsweisen ermöglichen.

B.7

On Covering the Gap between Computation and Humanities

Alexander Mehler, Andy Lücking - Goethe-Universität Frankfurt am Main

Dieser Beitrag wurde erstmals im Report des Dagstuhl Seminars 2014 veröffentlicht (http://drops.dagstuhl.de/opus/volltexte/2014/4792/pdf/dagrep_v004_i007_p080_s14301.pdf).

On Covering the Gap between Computation and Humanities

Alexander Mehler
Goethe University Frankfurt, DE
mehler@em.uni-frankfurt.de

Andy Lücking
Goethe University Frankfurt, DE
luecking@em.uni-frankfurt.de

October 23, 2014

Since digital or computational humanities (CH) has started it's triumph in the humanities' research landscape, it is advisable to have a closer look at its methodological and epistemological range. To this end, we look at CH from the point of view of preprocessing, machine learning, and the general philosophy of science and experimental methodology. From this perspectives, a number of gaps between CH on the one hand and classical humanities on the other hand can be identified. These gaps open up when considering: (i) the status of preprocessing in CH, its logical work-flow and the evaluation of its results compared to the needs and terminological munition of the humanities. Most importantly, corpus preprocessing often comes *before* hypothesis formation and respective model selection has been carried out, turning the logically as well as methodologically required workflow upside down. (ii) the predominant role of functional explanations in CH applications vs. the predominant role of intentional explanations with regard to the humanities. While computational processes can at most be functionally evaluated, hypotheses made in the humanities are usually embedded within contexts of justification that draw on some intentional statement. (iii) the possibilities of falsifying CH hypotheses and hypotheses in the humanities. Given the different typical patterns of explanations (see (ii) above), the results of computations and of the humanities cannot put to falsification as known as the powerful methodology from the natural, experimental sciences. This leaves open questions about the validity of these results. (iv) the use of big data in CH vs. the use of deep data in the humanities. Analyses in the humanities usually involve the interpretation and rational reconstruction of their objects. This hermeneutic procedure goes beyond mere preprocessing and

parsing of those objects, as is typically within reach of CH applications. When gathering interpreted and preprocessed data into corpora (which is done only seldom in the humanities, though), both approaches result in different kinds of resources which may be only of marginal benefit for the respectively other party. (vi) the lack of experimental methods in both CH and the humanities. In order to implement a notion of falsification in CH, one needs to think of CH-specific experimental settings which give rise to test procedures in the first place.

Based on these assessments, we argue that there are at least five interrelated gaps between computation and humanities, namely (1) an *epistemological gap* regarding the kind of evaluation mainly addressed by computational models in contrast to the kind of explanations addressed in the humanities; (2) a *data-related gap* regarding the build-up of ever growing text corpora in computer science in contrast to the need of controlled as well as deeply annotated data in the humanities; (3) a *semiotic gap* regarding signs as strings in the CH in contrast to rich sign-theoretical notions employed in the humanities; (4) a *methodological gap* with respect to understanding the functioning of methods of computer science by humanities scholars; and (5) an *interpretation gap* regarding the foundation of statistical findings in terms of the theoretical terms of the humanities involved. Having diagnosed these gaps we proceed by delineating two steps that could narrow (some of) these gaps: firstly, the understanding of CH technologies should be fostered by implementing them as part of a curriculum. Secondly, we should think of hybrid algorithmic methods, i.e. methods that at crucial branching points involve humanist expertise from the outset and in this way may pave the way towards "hermeneutic technologies" as a special kind of human-based evolutionary computing.

B.8

Informatik und Archäologie – Mein verflixtes 13. Jahr

Hubert Mara – FCGL – Forensic Computational Geometry Laboratory, IWR – Interdisciplinary Center for Scientific Computing, Heidelberg University

Informatik und Archäologie – Mein verflixtes 13. Jahr

Hubert Mara

FCGL – Forensic Computational Geometry Laboratory
IWR – Interdisciplinary Center for Scientific Computing, Heidelberg University
Im Neuenheimer Feld 368
69120 Heidelberg

Betrachtet man die Vielzahl an digitalen Messverfahren, die in der Archäologie zur Anwendung kommen, gibt es mehr als genügend archäologische Fragestellungen die zur Methodenentwicklung in der Informatik führen. Die Motivation wird besonders durch die Unschärfe der Fragen und die fehlende Grundwahrheit gefördert. Damit wird schnell klar, dass die Archäologie als Partnerdisziplin den gleichen technischen Stellenwert wie die Medizin hat, mit der ich in Industrieprojekten zusammengearbeitet habe. Mit dem Einwerben des ILATO Projekts konnten meine langjährige Kollegin und Mentorin Susanne Krömker und ich zeigen, dass Techniken, die an Keilschrifttafeln und jüdischen Grabsteinen entwickelt wurden auch zur Versuchsplanung in der industriellen Prüftechnik eingesetzt werden können.

Neben den technischen Herausforderungen existieren wesentlich schwerwiegendere Hindernisse im Zusammenfinden von Informatik und geisteswissenschaftlichen Fächern. Das beginnt bereits mit der Wahl des Studiengangs, da viele technikaffine Studienbeginner selten ein Fach der philosophischen Fakultät wählen. Es erschien mir auch ganz natürlich den eigenen Interessen und Fähigkeiten zu folgen und meine Ausbildung zum Ingenieur für Elektrotechnik durch ein Informatikstudium zu ergänzen. Damit kann ich – wie alle anderen Studienabgänger – ein klares Ausbildungsprofil vorweisen. Die klare Ausprägung einer Ausbildung, die wichtig für einen Berufseinstieg ist, ist ein gutes Argument gegen eine weitere Bindestrichinformatik. Letztere existieren zudem nahezu ausschließlich mit einem Bindestrich zu naturwissenschaftlichen Fächern, bei denen Arbeitsprozesse, Denkschemata, Publikationsfrequenz, -länge, -struktur, Projektdauer und mit vielen anderen Aspekten des wissenschaftlichen Alltags kompatibel sind. Insofern besteht die Notwendigkeit für brückenbildende Maßnahmen zu den Geisteswissenschaften.

Diese fehlende Brücke macht mein 13. Jahr verflixt, da ich eine Finanzierung zum Aufbau einer Arbeitsgruppe bekommen habe und mich auf der Suche nach zwei Informatikern mit einem Grundverständis für ein geisteswissenschaftliches Fach (oder umgekehrt) begeben habe. Aus dem vermeintlich naheliegendsten Bereich der *Digital Humanities* kam zu meiner Überraschung kein einziger Bewerber und auch sonst keinerlei Nachfrage. Mit Archäologen und -innen gab es einige Gespräche, die jedoch immer damit endeten, dass man Angst hat an der Vertiefung in der Informatik zu scheitern, obwohl Grundkenntnisse durch das Nebenfach Informatik bereits vorhanden waren. Aus den wenigen ernsthaften Bewerbungen aus der Informatik konnte ich bis dato ein Stipendium an einen Informatiker vergeben, der keine Scheu vor der exotischen Anwendung in Form von Keilschriftzeichenerkennung hat. Das zweite Stipendium wird vorraussichtlich zum Wintersemester vergeben. Es zeichnet sich mittlerweile ab, dass sich – aus diesem noch laufenden Verfahren – weitere interessante Erfahrungen rund um interdisziplinäre Promotionsvorhaben ergeben werden, die ich gerne zu einem späteren Zeitpunkt diskutieren kann.

In dieser verflixten Situation angekommen, musste ich mir die Frage stellen, wie diese entstanden ist. Den entscheidenden Hinweis auf die Antwort habe ich im Gespräch mit den Kollegen vom CCeH in Köln bekommen: meine Arbeit ist in nahezu gleichen Teilen in Publikationsmedien der Informatik und der Archäologie zu finden. Der Grundstein stammt aus meiner Zeit als Diplomand

durch die spannende *Hands- On* Möglichkeit in der Archäologie mit der Teilnahme an Grabungen in Israel und Peru mit dem 3D- Scanner der TU Wien. Dieser mehrwöchige Blick über die Schulter von Archäologen war der Schlüssel, um formulierbare Fragen zu identifizieren und gezielt Methoden z.B. für die Keramikanalyse aus der Numerischen Geometrie und der Mustererkennung umzusetzen. Einem Bauchgefühl folgend, habe ich diesen Ansatz zu Beginn meines Promotionsprojekts wiederholt und bin für ein halbes Jahr aus dem naturwissenschaftlichen Campus in ein Büro neben der Assyriologie gewechselt, um iterativ die Methodik zur besseren Lesbarkeit von Keilschrifttafeln voranzubringen. Nur in diesem sehr direkten persönlichen Kontakt ist der Brückenschlag zwischen sehr unterschiedlichen Fachgebieten gelungen, um ein digitales Werkzeug für Schrift in 3D zu entwicklen. Der umgekehrte Weg, dass z.B. ein Doktorand aus der Archäologie in einer Arbeitsgruppe der Robotik am IWR promoviert, existiert deutlich häufiger. Dieser Effekt ist wahrscheinlich – aber nicht ausschließlich – den unterschiedlichen Finanzierungslagen zuzuschreiben.

Aus meinen persönlichen Erfahrungen würde ich mir bei den geisteswissenschaftlichen Einrichtungen offenere Türen beim Zugang zu Objekten und Themen wünschen, um künftigen Informatikern den Zugang zur Kultur und Arbeitsweise der Geisteswissenschaften zu ermöglichen. Von meinen Kollegen in der Informatik erwarte ich mir einen umsichtigeren Umgang mit Versprechungen zu den Möglichkeiten der rechnergestützten Methoden. Im Bezug auf Lehre sehe ich vor allem Bedarf an einzelnen interdisziplinären Lehrveranstaltungen, die während des Studiums Einblicke und Anknüpfungspunkte zu technischen Möglichkeiten bzw. geisteswissenschaftlichen Fragestellungen bieten, um z.B. entsprechende Abschlussarbeiten anfertigen zu können.

B.9

Informatik – technischer Hilfsdienst der Digital Humanities?

Michael Piotrowski – Leibniz-Institut für Europäische Geschichte, Mainz

Informatik – technischer Hilfsdienst der Digital Humanities?

Dr.-Ing. Michael Piotrowski
Leibniz-Institut für Europäische Geschichte, Mainz

31. August 2014

Im Positionspapier »Was ist Informatik?« der GI (GI 2006, S. 16) heißt es:

> Nicht nur die Bedeutung der Informatik als Wissenschaft, auch ihre Rolle als Innovationsfaktor für andere wissenschaftliche Disziplinen ist inzwischen sprichwörtlich. Neben der Mathematik hat sich die Informatik in den letzten Jahrzehnten als weitere Querschnittswissenschaft etabliert. Durch Softwarewerkzeuge und Systeme, die höchst komplexe Berechnungen und Simulationen erlauben, wirkt die Informatik besonders nachhaltig und führt zu neuen Erkenntnissen und Methoden bis hin zu eigenen Teildisziplinen.

Handelt es sich bei den Digital Humanities um eine solche Teildisziplin? In den Geisteswissenschaften – und den Feuilletons – wird heftig über die Digital Humanities diskutiert: sind sie die Zukunft der Geisteswissenschaften oder vielmehr der »Ausverkauf« an die Informatik? Auch wenn mit der Gründung des DHd-Verbandes ein weitere Schritt in Richtung der Etablierung und Institutionalisierung der DH in Deutschland getan wurde, ist auch innerhalb den DH noch längst nicht klar, was die DH genau sind. Eine Frage, die dabei immer wieder aufkommt ist eben auch die, ob die DH eine eigene Disziplin *in den Geisteswissenschaften* sind.

In der Informatik – die ja quasi die andere Hälfte der DH darstellt – hat es bislang keine derartigen Diskussionen gegeben. Dies liegt sicher auch an einem Kulturunterschied: Auch wenn die Informatik durchaus zur Selbstreflexion fähig ist, muss sie sich nicht ständig ihrer Identität versichern. Es gibt daher auch keine Berührungsängste mit neuen Anwendungsgebieten. Ebenso wenig müssen sich innerhalb der Informatik verschiedene Forschungsgebiete zwingend voneinander abgrenzen – niemand würde etwa fordern, dass »Datenbanken« ein eigener Studiengang sein sollte. Und selbst wenn sich ein Feld – wie etwa die Computerlinguistik – (teilweise) institutionell verselbständigt hat, so ist das auch kein Problem: Computerlinguisten und Informatiker, die sich mit Sprachverarbeitung beschäftigen, treffen sich auf den gleichen Konferenzen. Insofern sind die Diskussionen in den Geisteswissenschaften aus Sicht der Informatik oft schwer nachvollziehbar.

Dennoch sollte sich die Informatik damit beschäftigen, und zwar sowohl aus wissenschaftlichen als auch aus wissenschaftspolitischen Gründen. Im eingangs erwähnten GI-Positionspapier heißt es weiter:

> Die Informatik spielt in anderen Wissenschaften zunächst häufig die Rolle einer Basistechnologie, mit der sich große Mengen von Informationen besser strukturieren, verarbeiten und aufbereiten lassen. Aber nahezu immer führt dieser Einsatz von Informatik zu eigenen Modellbildungen in den Anwendungsgebieten, so dass sich sehr schnell neue Fragestellungen, aber auch neue Erkenntnisse ergeben. Dadurch entstehen bisher

> nicht mögliche virtuelle Experimente, umfangreichere Analysen und neuartige Verfahren und Methoden der Untersuchung und Entwicklung. Eindrucksvolle Beispiele hierfür finden sich in der Biologie, Medizin, Chemie und Pharmazie, Meteorologie, Weltraumforschung und Astronomie, aber auch in den Künsten, der Architektur und den Wirtschaftswissenschaften. (GI 2006, S. 16)

Im Augenblick befinden sich die Geisteswissenschaften bzw. die DH weitgehend noch in dem Stadium, in dem die Informatik die Rolle einer »Basistechnologie« spielt. Meist werden in DH-Projekte selektiv einzelne Informatikwerkzeuge oder -methoden genutzt, etwa Datenbanksysteme, Visualisierungsverfahren oder Parser. Bei einer Definition der DH allein durch die Geisteswissenschaften besteht die Gefahr, dass die Informatik in dieser Rolle verharrt. Wenn sich das Feld jedoch weiterentwickelt – wie das auch bei den im Zitat genannten Fächern passiert ist –, wird die Übernahme fixfertiger Lösungen aus der Informatik nicht mehr reichen. Die Abkopplung von der Informatik kann dann dazu führen, dass in den DH aktuelle Forschungsergebnisse aus der Informatik nicht mehr wahrgenommen werden, das Rad neu erfunden und auf Synergien verzichtet wird – zum Schaden aller.

Ein weiteres mögliches Problem stellt sich im Bereich der Forschungsförderung. Wie z. B. die BMBF-Ausschreibungen zeigen, hat die Politik die Bedeutung der DH erkannt. Wenn aber die Geisteswissenschaften alleine definieren, was die DH sind und sich dabei – auch ungewollt – der Eindruck festsetzt, die Informatik werde nur für Programmieraufgaben benötigt, fehlt am Ende das Geld für genau die Informatikforschung, die eigentlich notwendig wäre, um die DH-Projekte realisieren zu können.

Schließlich entstehen in Deutschland jetzt explizit als »Digital Humanities« deklarierte Studiengänge. Auch hier sollte sich die Informatik einbringen, um beispielsweise zu vermeiden, dass sich der Informatikanteil solcher Studiengänge allein aufs »Programmieren« beschränkt, was letztlich aber nicht ausreichen wird.

Im Feuilleton der FAZ konnte man vor einiger Zeit staunend lesen:

> Das Quellgebiet der digitalen Geisteswissenschaften sind Computerlinguistik, Computerphilologie und die Fachinformatiken. Man kann in ihnen so etwas wie den technischen Hilfsdienst der Geisteswissenschaften sehen, dem bisher die wissenschaftliche Anerkennung ihrer Leistungen und das Verständnis und die Rückendeckung ihrer Fächer versagt blieb. (Thiel 2012)

Wenn die Informatik nicht nur als »technischer Hilfsdienst« verstanden werden will, sondern auch selbst Gewinn aus der Zusammenarbeit mit den Geisteswissenschaften ziehen will, dann darf sie die Deutungshoheit nicht allein den Geisteswissenschaftern überlassen. Daher ist es meines Erachtens notwendig, dass die Informatik – z. B. über ihre Fachgesellschaften – das Feld »Informatik in den Geisteswissenschaften« selbst besetzt. Anzumerken wäre hier noch, dass in der Computerlinguistik die GSCL seit 2013 einen Arbeitskreis »Digital Humanities« hat und die ACL bereits seit 2012 die SIGHUM.

Literatur

GI, Hrsg. (2006). *Was ist Informatik?* Positionspapier der Gesellschaft für Informatik. URL: `http://www.gi.de/fileadmin/redaktion/Download/was-ist-informatik-lang.pdf`.

Thiel, Thomas (2012). *Digital Humanities. Eine empirische Wende für die Geisteswissenschaften?* FAZ.NET. URL: `http://www.faz.net/-gsn-71khe`.

B.10

Auf dem Weg zu einer engeren Verzahnung von Informatik und Geisteswissenschaften

Caroline Sporleder, Ralph Bergmann, Thomas Burch, Stephan Diehl, Vera Hildenbrandt, Claudine Moulin, Ingo J. Timm – Universität Trier

Auf dem Weg zu einer engeren Verzahnung von Informatik und Geisteswissenschaften

Caroline Sporleder, Ralph Bergmann, Thomas Burch, Stephan Diehl, Vera Hildenbrandt, Claudine Moulin, Ingo J. Timm

Universität Trier

In den letzten Jahren hat das Interesse an der Verarbeitung geisteswissenschaftlicher Daten in den informatiknahen Fächern mehr und mehr zugenommen. So hat zum Beispiel die Computerlinguistik in den Geisteswissenschaften eine reiche Datenquelle für die Anwendung und Weiterentwicklung robuster Sprachverarbeitungsalgorithmen entdeckt. Geisteswissenschaftliche Daten sind gerade deshalb für die Computerlinguistik interessant, weil sie oft Eigenschaften haben, die sie "schwieriger" machen als die gängigen Benchmark-Datensets, die meist aus modernen, born-digital und umsichtig editierten Zeitungstexten bestehen. Aus der Hinwendung zu neuen geisteswissenschaftlichen Daten ergeben sich daher völlig neue Forschungsfragen für die Computerlinguistik, z.B. hinsichtlich der Verarbeitung älterer Texte mit uneinheitlicher Orthographie und abweichender Grammatik. Genauso gibt es in den nicht-sprachverarbeitenden Bereichen der Informatik zahlreiche Aufgabenbereiche an der Schnittstelle zu den Geisteswissenschaften, die die Entwicklung völlig neuer Verfahren oder zumindest eine Weiterentwicklung bestehender Methoden verlangen, z.B. spezielle Bildverarbeitungsverfahren zur Rekonstruktion antiker Fundstücke. Auf der anderen Seite existieren unzweifelhaft auch wichtige Digital-Humanities-Projekte, in denen der informatische Aspekt nur die Anwendung, Implementierung und gegebenenfalls leichte Adaption bestehender Techniken betrifft, wie es u.a. bei vielen Digitalisierungsprojekten der Fall ist.

Langfristig wirklich vielversprechend und für die Informatikwissenschaften interessant ist die Zusammenarbeit mit den Geisteswissenschaften aber nur dann, wenn sich neben den geisteswissenschaftlichen auch innovative informatikwissenschaftliche Forschungsfragen ergeben. Nur in einem solchen Rahmen ist es möglich, genuin informatikwissenschaftliche Forschung zu betreiben und entsprechende Doktorandenstellen zu schaffen. Projekte mit einem hohen Innovationspotenzial in beiden Bereichen stellen die Beteiligten jedoch vor eine Reihe von Herausforderungen. Neben der sehr zentralen Frage, wie der –in der Regel– hermeneutische Forschungsprozess in den Geisteswissenschaften durch informatikwissenschaftliche Techniken unterstützt werden kann, ergeben sich auch eine Reihe von ganz praktischen Problemen:

Asynchronizität In einem typischen DH-Projekt mit Innovationspotenzial in beiden Bereichen ist es oft schwierig geistes- und informatikwissenschaftliche Forschung vollkommen synchron zu betreiben. Die beteiligten Geisteswissenschaftler sind für ihre Forschung typischerweise auf Software angewiesen, die im Projektrahmen noch entwickelt werden muss. Diese Software ist aber das Ergebnis der Beantwortung einer informatikwissenschaftlichen Forschungsfrage, steht also aus informatikwissenschaft-

licher Sicht am Projektende. Dieses Problem kann durch iterative Prototypenentwicklung und -anwendung teilweise entschärft werden. Ein ähnliches, gegenläufiges Problem ergibt sich, wenn Informatikwissenschaftler für ihre Forschung auf geisteswissenschaftliche Daten angewiesen sind, die im Projektrahmen zuerst noch erschlossen bzw. aufbereitet werden müssen.

Unterschiedliche Publikationsanforderungen Die Publikationsanforderungen in beiden Bereichen unterscheiden sich zum Teil deutlich (Zeitpunkt, Publikationsform, Anzahl der Autoren, Reihenfolge der Autoren). Gemeinsame Publikationen bieten daher ein gewisses Konfliktpotenzial, wenn nicht spätestens zu Beginn des Projekts konkrete Absprachen getroffen werden.

Quantifizierbarkeit und Erfolgskriterien Geisteswissenschaftliche Forschung ist oft eher qualitativ geprägt, während informatikwissenschaftliche Forschung meist quantitativ ist. Geisteswissenschaftler können daher zum einen quantitative, probabilistische Ergebnisse manchmal nicht korrekt einordnen. Zum anderen sind manche Ergebnisse, die von einem informatikwissenschaftlichen, quantitativen Blickpunkt aus akzeptabel sind, vom geisteswissenschaftlichen Standpunkt aus ungenügend (z.B. Fehlertoleranzen).

An der Universität Trier verfolgen wir seit langem eine enge Verzahnung von Informatik- und Geisteswissenschaften in Forschung und Lehre mit dem Ziel, Fragestellungen mit Innovationspotenzial für beide Bereiche zu identifizieren und darüber hinaus durch eine enge Zusammenarbeit das gegenseitige Verständnis zu vergrößern und die oben genannten Problemfelder zu vermeiden. In der Lehre bieten die Fächer Informatik und Computerlinguistik & Digital Humanities zusammen mit dem Trier Center for Digital Humanities einen gemeinsam entwickelten Masterstudiengang “Digital Humanities” an, der informatikwissenschaftliche Kompetenzen gleichberechtigt neben geisteswissenschaftliche und solche im Kernbereich der Digital Humanities stellt. Neben gemeinsam betreuten Abschlussarbeiten haben wir auch ein (universitätsinternes) Austauschprogramm etabliert, das es Nachwuchswissenschaftlern aus den Informatikwissenschaften ermöglicht, an DH-Forschungsprojekten mitzuarbeiten. Im Bereich der Forschung liegt der Schwerpunkt auf der Entwicklung innovativer und generisch-modular anwendbarer Konzepte, Methoden und Werkzeuge. Dies umfasst nicht nur bereits gut erprobte, eher statische Verfahren der Ergebnisvisualisierung, sondern betont vor allem den prozessualen Charakter von Datenauswertung, Deutung und Ergebnisfindung sowie den interaktiven Datenzugang durch Simulation.

Teil C

Beiträge der Organisatoren

C.1

Anmerkungen zum Verhältnis von Digital Humanities und Informatik

Günther Görz – Friedrich-Alexander Universität Erlangen-Nürnberg

Anmerkungen zum Verhältnis von Digital Humanities und Informatik

Günther Görz
FAU Erlangen-Nürnberg, Dept. Informatik und MPIWG, Berlin

These

„Digital Humanities“ steht in einem analogen Verhältnis zu den Geisteswissenschaften wie „Computational Science“ zu den Naturwissenschaften bzw. „Computational Engineering“ zu den Ingenieurwissenschaften.

Es geht an erster Stelle um eine Ergänzung im Methodeninventar. Sicher ist: Neue Größenordnungen in den Daten erfordern neue Auswertungsverfahren. Um aber bei der Analogie zu bleiben: Was entscheidend hinzukommt, sind Modellbildung und Simulation. Ob auch epistemologisch etwas Neues entstehen wird, bleibt abzuwarten; vielleicht liegt es in einer Erweiterung der Epistemologie um eine algorithmische und experimentelle Dimension.

Auf jeden Fall wäre es verfehlt, in den Digital Humanities die Aufkunft einer neuen Wissenschaft zu sehen — man bedenke nur, was aus den Ansprüchen der vermeintlichen Superwissenschaft Kybernetik aus den 1950er Jahren geworden ist.

Dass andererseits die Informatik dabei die Rolle einer reinen Hilfswissenschaft spielen muss, ist noch lange nicht gesagt. Lehrreich mag hier sein, auf das Verhältnis von Computerlinguistik und Sprachverarbeitung („Natural Language Processing“) zu schauen: Die Sprachverarbeitung, die in den „Digital Humanities“ selbst zahlreiche Anwendungen findet, hat durchaus eine eigene Akzentuierung. Sie unterscheidet sich von der Computerlinguistik vor allem im Erkenntnisinteresse und in der Anwendung, auch wenn sich beide der gleichen algorithmischen Mittel bedienen: Die Sprachverarbeitung hat kein primäres Interesse an grammatischen Strukturen aus linguistischer Sicht, sondern sieht sie als eine Zwischenrepräsentation auf dem Weg zur semantisch-pragmatischen Inhaltsanalyse — mitgeteilte Bedeutungen in gesellschaftlichen Kommunikationsprozessen, die in symbolischer Form repräsentiert werden. Das schließt den Einsatz stochastischer Methoden nicht aus, aber letztlich geht es nicht um Wahrscheinlichkeitsverteilungen, sondern um in formalsprachlicher Form repräsentierte kommunikative Inhalte.

Zur Erklärung

Ersichtlich kann an dieser Stelle kein Definitionsversuch der „Geisteswissenschaften“ erfolgen. Stattdessen gehen wir einfach von dem durchaus heterogenen Fächerkanon aus, der an den philosophischen Fakultäten unserer Universitäten trotz aller neoliberalen Demontageversuche immer noch vorhanden ist. Gemeinsam ist ihnen ihr Selbstverständnis als Orientierungswissenschaften: Es geht um das Verstehen von Handlungen, um Erklären, Rekonstruieren und Planen.

Die fundamentale Abstraktionsleistung der Informatik kann man in der Schematisierung und Automatisierung von Verstandesleistungen sehen. In historischer Perspektive gibt es eine entsprechende Tradition, spätestens mit Leibniz beginnend, über Babbage, Turing und

Zuse — um nur einige der Informatik-Pioniere zu nennen.

Generell richtet sich an die Informatik die Frage, welche Beiträge ihre Forschungsansätze und Methoden, etwa zur Wissensrepräsentation und Inferenz, zur informierten Suche, zur Textanalyse und semantischen Erschließung, zur Bildanalyse und Objekterkennung, zur Visualisierung komplexer Datenstrukturen, zur Planung und Problemlösung zu geisteswissenschaftlichen Erklärungen liefern können.

Eine echte Wechselwirkung zwischen Geisteswissenschaften und Informatik gibt es bisher nur in eher bescheidenem Umfang. Zum einen kann die Informatik lernen, in einen rationalen Diskurs über Wissensformen einzutreten. Dabei hat sie aus langer Beschäftigung mit Wissensrepräsentation und -verarbeitung auch methodisch etwas zu bieten, z.B. kann sie bei der Bereitstellung eines sprachlichen Rahmens zur begrifflichen Modellierung und Begründung helfen und Hilfsmittel zur Klassifikation, Annotation oder Verknüpfung von Wissenselementen beitragen.

Mit der von formalen Mitteln geprägten Erforschung symbolischer Strukturen, wie sie uns in Antike, Mittelalter, früher Neuzeit und in außereuropäischen Kulturen begegnen, ist weitgehend Neuland zu betreten. Begriffssysteme haben ihre ihnen eigene historische Dynamik: Es geht um die Entwicklung und den Einsatz von Standards für Namen, Bezeichnungen, Fachtermini und ihre Systematisierung in Thesauri und formalen Ontologien, die zugleich ihren — durch vielfältige Transferbeziehungen beeinflußten — historischen Wandel mitbedenken.

Kognitionswissenschaftliche Aspekte

Angesichts der in der Interaktion der Disziplinen relevanten kognitiven Leistungen wie Wahrnehmung, Gedächtnis, Analyse und Synthese wird es für die Kognitionswissenschaft, sofern sie sich nicht nur naturwissenschaftlich-reduktionistisch versteht — also Handlungsverstehen im Sinne der Hermeneutik thematisiert —, problematisch mit Ansätzen zur Operationalisierung: Wenn man Handeln auf Verhalten reduziert, ist man gleich wieder bei den Naturwissenschaften. Naturalistische Hermeneutik ist eine contradictio in adjecto. Die ganze Problematik wird deutlich an der Diskussion um die Rolle bzw. den Erkenntnisgewinn bildgebender Verfahren: Sie zeigen nur Korrelationen, aber keine Kausalität — auch wenn Letzteres von wissenschaftstheoretisch wenig gebildeten Vertretern der Zunft immer wieder frech, aber medienwirksam behauptet wird.

Stattdessen möchte ich vorschlagen, eine fruchtbringende Synthese in einer Epistemologisierung kognitionswissenschaftlicher Ansätze zu suchen, d.h. in einer Reinterpretation auf der Wissensebene: Welche Wissensarten sind denn im Spiel, wenn bestimmte kognitionswissenschaftliche Modelle konstruiert werden, und von welchem Wissenstyp sind diese Modelle selbst? Das ginge in die Richtung einer epistemischen Anreicherung des "Linguistic Turn" unter Berücksichtigung der historischen Dimension. Nelson Goodman hat das einmal so ausgedrückt, dass die Struktur des Geistes — also der Mentalismus — abgelöst wird durch die Struktur der Begriffe und diese wiederum durch die verschiedenen Symbolsysteme der Wissenschaften und der Philosophie. Damit sich nicht alles in postmoderner Beliebigkeit auflöst, ist auf die Einheit der wissenschaftlichen Rationalität zu verweisen. Diese wird praktisch garantiert durch die Verfahren der Verifikation und der bedeutungskonstituierenden Begründung in der (Wissenschafts-) Sprache.

Auch was unser Verständnis von Inferenz betrifft, wird sich einiges ändern: Neben die Explikation impliziten Wissens tritt die Verarbeitung von Massendaten v.a. mit statistischen Verfahren, was mit deren Verfügbarkeit durch das Internet und grid-förmig organisierter massierter Verarbeitungskapazität erst seit der Mitte der neunziger Jahre möglich geworden ist. So wird mittelfristig der Einsatz der digitalen Techniken wohl auch zu einer Veränderung der Forschungsmethoden und -strategien in den jeweiligen geisteswissenschaftlichen Disziplinen führen, so dass hier eine echte Wechselwirkung zu erwarten ist.

C.2

Informatik und die Humanities

Andreas Henrich – Otto-Friedrich-Universität Bamberg

Informatik und die Humanities

Ich habe im Titel ganz bewusst das „Digital“ der so genannten „Digital Humanities“ weggelassen. Dazu werde ich im Folgenden eine pointierte Position beziehen. Natürlich gibt es viele Graustufen, aber manchmal ist es auch sinnvoll eine Position bei aller Angreifbarkeit in schwarz-weiß zu zeichnen.

Insgesamt sind aus meiner Sicht drei Fragen miteinander verbunden:

Werden die Humanities zu Digital Humanities?

Der Gegenstand und die Fragen der Geisteswissenschaften bilden hier das Zentrum. Ob diese mit digitalen oder analogen Methoden angegangen werden, ist dabei (zumindest auf lange Sicht) nicht zentral. Genauso wie in anderen Disziplinen (Ingenieurswissenschaften, Biologie, BWL …) werden digitale Methoden langfristig zum selbstverständlichen Repertoire des Geisteswissenschaftlers gehören. Der eine wird sie öfter nutzen als der andere, aber die Hervorhebung des „Digital“ mag in einer kurzen Phase der Erweiterung der Möglichkeiten sinnvoll sein, langfristig sehe ich dies aber nicht. Hinzu kommt, dass ich zahlreiche Geisteswissenschaftler kennengelernt habe, die ohne Berührungsängste digitale Verfahren nutzen, sich aber keineswegs als „Digital Humanist“ bezeichnen oder verstehen würden.

Eine wichtige Frage wird dabei sein, ob die digitalen Methoden über die Geisteswissenschaften hinweg zumindest zu einem hinreichenden Teil so generisch sind, dass die übergreifende Betrachtung wirklich zielführend ist. Wie groß sind da die Überlappungen z. B. zwischen den Philologien und materiellen Kulturwissenschaften? Würden „Digital Humanities“ nicht ohnehin allenfalls ein (recht vager) Oberbegriff sein?

Wie wird die Digitalisierung der Geisteswissenschaften die Informatik verändern?

Ja: die Digitalisierung der Geisteswissenschaften wird die Informatik verändern! Aber nicht mehr und nicht weniger als die Wirtschaftsinformatik, die Bioinformatik …

Natürlich gibt es (viele) spannende Probleme in den Geisteswissenschaften, die die Forschung in der Informatik vor Herausforderungen stellen. Dies unterscheidet sich aber aus meiner Sicht nicht grundsätzlich von anderen Anwendungsgebieten.

Wird es eine neue Bindestrich-Disziplin geben?

Eigentlich müsste man hier weiter differenzieren. Zunächst ist die Frage, wie die beiden verbundenen Disziplinen von der Ausrichtung und vom Gewicht her zueinander stehen. Natürlich kann man Geisteswissenschaftler mit digitalen Methoden in ihrer Arbeit unterstützen (und diesen Fall möchte ich hier adressieren). Eine andere durchaus auch sinnvolle aber komplett andere Betrachtung wäre eine geisteswissenschaftliche (oder z.B. soziologische) Perspektive auf die Informatik, ihr Schaffen und ihre Auswirkungen. Wenn wir also hier einmal von einer Bindestrich-Informatik ausgehen, dann ist die nächste Frage, „wo“ diese anzusiedeln ist. Auf diese Frage gibt es aus meiner Sicht keine richtige oder falsche Antwort. Die Antwort muss vielmehr aus dem Kontext und auf Basis von Überlegungen der Zweckmäßigkeit gegeben werden. Am Beispiel der Wirtschaftsinformatik (WI) kann man sehen, dass es Standorte gibt wo diese als spezielle BWL gesehen wird und andere Standorte, wo die WI in der Informatik liegt. Viel wichtiger als diese Zuordnung erscheint mir, dass diejenigen, die die Bindestrich-Disziplin vertreten, beide Gebiete in der notwendigen Breite und Tiefe verstehen – und im Idealfall hierzu noch eine originäre dritte „Säule“, die den Erkenntnisgegenstand und den speziellen

methodischen „Werkzeugkasten“ zum Inhalt hat. Hier kann die Wirtschaftsinformatik dann durchaus als Vorbild dienen.

Eine andere Bedeutung bekommt die Zuordnung natürlich bei Studiengängen. Auch hier gibt es aus meiner Sicht keine richtige oder falsche Zuordnung. Wichtig ist, dass die Studiengangsbezeichnungen deutlich machen, ob es sich um einen Studiengang der Angewandten Informatik oder um einen Studiengang der Geisteswissenschaften mit einem Fokus auf digitalen Methoden handelt.

Konsequenzen für die Informatik

Die Informatik sollte aus meiner Sicht aus diesen Überlegungen folgende Konsequenzen ziehen:

- Man sollte intensiv mit Wissenschaftlerinnen und Wissenschaftlern in den Geisteswissenschaften kooperieren, die offen für die innovative Nutzung digitaler Verfahren für Ihre Fragestellungen sind – und zwar unabhängig davon, ob sich diese als „digitale Geisteswissenschaftler“ oder schlicht als „Geisteswissenschaftler“ verstehen.
- Die Informatik muss selbst Stellung beziehen zu Fragen der Digitalisierung der Geisteswissenschaften. Dies reicht von curricularen Empfehlungen bis zu einer Repräsentation in den Fachgliederungen der GI. Es geht dabei nicht darum, mit den Geisteswissenschaften um die Hoheit über das Thema „Digital Humanities“ zu kämpfen. Vielmehr geht es darum, neben den durch den Verband DHd („Digital Humanities im deutschsprachigen Raum“) repräsentierten Blickwinkel einen informatischen Blickwinkel zu setzen – unter anderem, um hier gleichberechtigt in einen Diskurs treten zu können.
- Schließlich sollte auch darauf hingewirkt werden, dass die informatische Perspektive in den Berufungskommissionen zu den nun vielerorts ausgeschriebenen Professuren im Feld der „Digital Humanities“ hinreichend (was auch immer dies bedeuten mag) vertreten ist. An dieser Stelle werden aktuell Weichen gestellt, die einzeln und vor allem in Summe ganz entscheidend für die Entwicklung sein werden.

C.3

Digital und Computational Humanities

Gerhard Heyer, Daniel Isemann –
Universität Leipzig

Digital und Computational Humanities

Gerhard Heyer, Daniel Isemann
Universität Leipzig

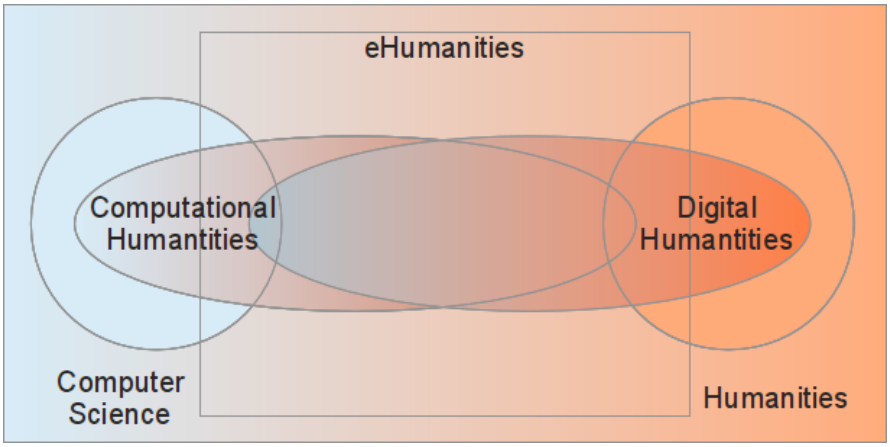
Während Informatik und Geisteswissenschaften bisher in ihren Arbeitsmethoden oft eher als Gegenpole aufgetreten sind, anstatt sich auf mögliche Synergien zu konzentrieren, betreten wir mit dem Aufkommen der digitalen Geisteswissenschaften einen neuen Bereich der Interaktion zwischen den beiden Bereichen. Für die Geisteswissenschaften ist dies die Verwendung von computergestützten Methoden, die eine effizientere Forschung und neue Forschungsfragen ermöglichen, die ohne solche Methoden nicht hätten behandelt werden können. Für die Informatik kann die Hinwendung zu den Geisteswissenschaften als einem Anwendungsgebiet ebenso neue Problemstellungen ergeben und dazu führen, vorhandene Ansätze zu überdenken und die Entwicklung neuer Lösungen zu begünstigen, die helfen können die Informatik auch in anderen Bereichen medienorientierter Anwendungen voranzutreiben. Aber die meisten dieser Lösungen bleiben derzeit auf einzelne Projekte beschränkt und erlauben der wissenschaftlichen Gemeinschaft in den digitalen Geisteswissenschaften nicht, von Fortschritten in anderen Bereichen der Informatik wie Visual Analytics zu profitieren. Dies wird erst möglich, wenn wir über einzelne (vertikale) Anwendungen in den Geisteswissenschaften hinausgehen und eine fächer- und projektübergreifende (horizontale) Perspektive einnehmen.

In der Folge daraus ist es wichtig, zwei Arten oder Aspekte „digitalen" Arbeitens in den Geisteswissenschaften zu unterscheiden:

(1) die Schaffung, Verbreitung und Nutzung von Digitalisaten bestehender Inhalte und von Repositorien solcher Digitalisate, und

(2) die computergestützte Analyse und Aufbereitung von digitalen Repositorien mit Hilfe modernster algorithmischer und informatischer Methoden.

Während die erste ursprünglich aus den Geisteswissenschaften selbst heraus entstanden ist und häufig als *Digital Humanities* bezeichnet wird, impliziert die zweite eine Stärkere Präsenz computergestützter Ansätze und könnte daher vielleicht als *Computational Humanities* bezeichnet werden.



Figur 1. Positionierung der Computational und Digital Humanities im Kontext von Informatik (Computer Science) und Geisteswissenschaften (Humanities).

Zwischen diesen beiden Aspekten zu unterscheiden hat erhebliche Auswirkungen auf die tatsächliche Arbeit, die durchgeführt wird. Im Hinblick auf das Know-how der Forscher und ihrer institutionellen Verortung entweder in den Geisteswissenschaften oder in der Informatik, kann

ihre Forschung entweder mehr auf die Schaffung und Nutzung von Digitalisaten und entsprechenden Repositorien oder auf die Entwicklung und Umsetzung genuin programmatischer und konzeptueller Lösungen für die Geisteswissenschaften als einem Bereich der angewandten Informatik abzielen.

Die Bündelung von computergestützten Methoden für die Geisteswissenschaften in einem Bereich *Computational Humanities* scheint uns, gegenüber der individuellen Verortung solcher Bestrebungen in den Einzeldisziplienen, drei Vorzüge zu bieten: Erstens, ein verbesserter und flexiblerer Zugang der Geisteswissenschaften zu algorithmischen Methoden die generisch genug sind, um fachübergreifend Anwendung zu finden, darunter auch solchen die sich auf jeweils jüngste und neueste Forschungsergebnisse stützen. Solche Methoden, die die Informatik für verschiedene geisteswissenschaftliche Kontexte anbieten kann betreffen, unter anderem,

- die qualitative Verbesserung digitaler Quellen (Standardisierung von Schreibvarianten, Rechtschreibkorrektur, die eindeutige Identifizierung von Autoren und Quellen, die Kennzeichnung von Zitaten und Referenzen, die algorithmengestützte zeitliche Einordnung der Texte, etc.);
- die Menge und Struktur der Quellen, die skalierbar verarbeitet werden können (Verarbeitung von sehr großen Textmengen, Strukturierung nach Zeit, Ort, Autoren, Inhalten und Themen, Anmerkungen von Kollegen, Kommentaren aus anderen Ausgaben, etc.);
- die Art und Qualität der Analyse (breite, datengetriebene Studien, strikter Bottom-up-Ansatz unter Verwendung von Text-Mining-Tools, Integration von Community Networking Ansätzen, Kontextualisierung von Daten, etc.).

Zweitens können zusätzlich zu diesen generischen algorithmischen Methoden unter Verwendung von Erkenntnissen aus den Bereichen der Informationswissenschaft, der computergestützten Ontologien und Wissensdatenbanken digitale Ressourcen (wie zum Beispiel Geo- oder Personendatenbanken) aufgebaut werden, die sich disziplinenübergreifend wiederverwenden lassen, da sie den Anforderungen nicht nur eines, sondern mehrerer Fachgebiete Rechnung tragen, und flexibel genug gestaltet sind, um auch auf künftige Veränderungen im Anforderungsprofil reagieren zu können.

Drittens und letztens könnte die Ausbildung einer quasi-permanenten horizontalen Schicht der *Computational Humanities* zwischen den Säulen (oder Silos) der einzelnen Geisteswissenschaften zu einer gegenseitigen methodischen Befruchtung unter den geisteswissenschaftlichen Disziplinen führen. Fachspezifische, nicht-informatische Methoden könnten unter der Linse der Informatik betrachtet Ähnlichkeiten oder Verschiedenheiten aufweisen die einen methodologischen Dialog zwischen den geisteswissenschaftlichen Disziplinen anregen können. Dieser letzte Punkt ist zugegebenermaßen der Spekulativste, aber für die Geisteswissenschaften vielleicht der Wirkmächtigste.

Die Genese neuer wissenschaftlicher Betätigungsfelder in ihrem historischen und ideengeschichtlichen Umfeld gibt nicht notwendigerweise den besten Aufschluss darüber, in welche Richtung sich diese entwickeln werden. So wurde die moderne Atomtheorie von einem gelernten Meteorologen (Dalton) begründet, und der Erfinder des Begriffs Sauerstoff (Lavoisier) verband damit Ideen, die wir heute sicher ablehnen würden (u.a. dass es sich dabei um den zentralen Bestandteil von Säuren handelt, daher *Sauer*stoff). Die Tatsache, dass zahlreiche Digital Humanities Projekte institutionell und strukturell in einzelnen Geisteswissenschaften angesiedelt sind, muss nicht bedeuten, dass dies der einzige oder auch nur der beste Ansatz ist, um informationswissenschaftliche oder computergestützte Methoden für die geisteswissenschaftliche Forschung fruchtbar zu machen. Das skizzierte Betätigungsfeld der Computational Humanities lässt sich als Ergänzung der bestehenden Digital Humanities begreifen, das generische Methoden der Informatik über traditionelle geisteswissenschaftliche Fachgrenzen hinweg unter Umständen besser zum Einsatz bringen kann. Darüberhinaus könnte es zum Aufbau von digitalen Ressourcen beitragen, die über bloße Digitalisate weit hinausgehen und von mehreren Geisteswissenschaften gemeinschaftlich und nachhaltig genutzt werden können. Am Ende können die Computational Humanities vielleicht sogar als Katalysator für Synergien oder präzisere Abgrenzungen zwischen ureigentlich geisteswissenschaftlichen Methoden fungieren.

C.4

What distinguishes Humanities' texts from other data?

Manfred Thaller – Universität zu Köln

What distinguishes Humanities' texts from other data?

Manfred Thaller, University at Cologne

(i) The Humanities are a very broad field. The following ideas relate to those Humanities disciplines, which are dealing with "historical texts" –or at least they started from them. "Historical" in this context defines any text, which has been created by actors, which we cannot consult any more. This creates a complication when we understand an existing text as a message from a sender to a recipient – an understanding which is absolutely fundamental to modern information technology, as it is the model which has been used within Shannon's article of 1948, one of the corner stones of modern information theory and for most computer scientist, *the* corner stone of Computer Science upon which the later has been built. All of the measures Shannon proposes require an understanding, what the message that has been transmitted by the sender contained before transmission. Another important restriction Shannon acknowledges himself:

> *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have* meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.*
> (Shannon, 1948, 379)

The fact that information processing systems start with a model which ignores semantics from page one. is ultimately the reason, why meaning has to be added to the signal stream in ways, which allow the transmission (or processing) of that information as an integral part of the signal stream – today usually as embedded markup. Embedded into a signal stream, which has been created by a sender; so embedding anything into it would, according to the model of Shannon, require the markup being part of the transmitted message. This is indeed, what SGML has been created for: To enter the intentions of the producer of a document about the formatting (and, up to a degree the meaning) of a data stream in such a way, that they would be independent of the requirements of specific devices.

When we are not able to check the meaning of a message with the sender we have to distinguish between the message, even if we do not understand it, and our assumptions about interpreting them. As we do not know the intent of the sender, the result of the "transmission"of a historical text across time cannot be determined conclusively.

(ii) That data – as transmitted in signal streams – and information, as handled by humans, are not identical is a truism. They have long been seen as separate strata in information theory. (For a recent overview of the discussion see Rowley 2007.) A main difference between Shannon and the "data – information – knowledge – wisdom" hierarchy has always been, that the former leads directly to an intuitive understanding of systems which can be realized by software engineering, while the later cannot. This is also true of attempts to use a similar scheme to understand information systems, notably Langefors 1995 *infological equation*.

(1) $I = i\ (D,\ S,\ t)$

Information (I) is understood here as the result of a process of interpretation (i) that is applied to data (D), applying previous knowledge (S) within the time available (t). The great attraction of this model is that – unlike Shannon's – it explicitly promises to model the meaning of messages, which are explicitly excluded from consideration by Shannon. To

emphasize the difference between the models, we could say that Shannon assumes information to exist *statically*, therefore it can be broken into discrete units, independent of any process dealing with it, while Langefors understands information to be the result of a *dynamic* process, which, having a relationship to time, goes through different stages: So the amount of information existing at t_n is not – or not necessarily – equal to the amount of information at t_{n-1}, the ongoing process i having had the chance to produce more of it in the meantime.

The previous knowledge – S – can of course be easily seen as embodied in the interpreting scholar, who looks at the data. For the creation of systems of information processing Thaller (2009a: 228) has shown that Langefors original equation can be developed further. When we assume that knowledge is transformed from a static entity into a dynamic process, as Langefors has proposed for information, we can – via a few steps omitted in this abstract - reach

(2) $I_x = i\ (I_{x-\alpha},\ s(I_{x-\beta},\ t),\ t)$

Roughly: Information at point *x* is the result of the interpretation of an earlier level of information, in the light of knowledge generated from earlier knowledge, at a point of time *t*. As this allows the interpretation of data – e.g. a "transmission" of a sender not living any more - as a process, which does not have to terminate, it is a better model for the handling of Humanities' texts as Shannon's.

(iii) This abstract model can be turned into an architecture for a representation of information, which can be processed by software. Thaller (2009b) has lead a project team within the digital preservation project PLANETS (cf. http://www.planets-project.eu/), which used this abstract model for the development of tools, which work on the comparison of the information contained within two different representations of an item according to two different technical formats. (Roughly: Does a PDF document contain exactly the same "text" as a Word document.) For this purpose it is assumed, that all information represented in persistent form on a computer consists of a set of tokens carrying information, which exists within an *n*-dimensional interpretative space, each dimension of that space describing one "meaning" to be assigned to it. Such a meaning can be a request directed at the rendering system processing the data to render a byte sequence in a specific way, or a connection to a semantic label empowering an information retrieval system. As such a representation is fully recursive, the requirements of formalism (2) above are fulfilled. For texts this can be simplified to an introductory example, where a text is seen as a chain of characters, each of which can be described by arbitrarily many *orthogonal* properties. (Whether the string *Biggin* within a text describes a person or an airfield is independent of whether that string is represented as italics or not; whether the string "To be or not to be" is assigned to the speaker *Hamlet* is independent of whether it appears on page 13 or 367 of a book.)

(iv) Returning to the argument of section (i) we can see, that there is a direct correspondence between the two arguments. On the one hand the necessity to keep (a) the symbols transmitted within a "message" from a sender who is irrevocably in the past and (b) our intellectual interpretations of them cleanly and unmistakably separate. On the other hand the necessity to distinguish clearly between (a) the tokens which transmit the data contained within a byte stream and (b) the technical information necessary to interpret that byte stream within a rendering system. If it is useful to transfer information transported within files with different formats into a representation, where the transmitted data are kept completely separate from the technical data needed to interpret them on a technical level, it is highly plausible, that that

is even more the case, when we are discussing interpretations of texts left to us by authors we can not consult any more.

This in turn is highly compatible to an architecture for virtual research environments for manuscript related work, where Humanities' work on historical texts is understood to consist of adding layers of changing and potentially conflicting interpretation unto a set of images of the manuscript to be interpreted. Ebner et al. 2011 have recently described an architecture for a virtual research environment for medieval manuscripts which implements this overall architecture, though using embedded markup for some of the layers for the time being.

To summarize the argument: (1) All texts, for which we cannot consult the producer, should be understood as a sequence of tokens, where we should keep the representation of the tokens and the representation of our interpretation thereof completely separate. (2) Such representations can be grounded in information theory. (3) These representations are useful as blueprints for software on highly divergent levels of abstraction.

References

Ebner, Daniel; Graf, Jochen; Thaller, Manfred (2011): "A Virtual Research Environment for the handling of medieval charters", paper presented at the conference *Supporting Digital Humanities: Answering the unaskable*, Copenhagen November 17th / 18th, 2011. (http://www.hki.uni-koeln.de/sites/all/files/VdUSDH11-02.pdf)

Langefors, Börje (1995): *Essays on Infology*, Lund: Studentliteratur.

Rowley, Jennifer (2007): "The wisdom hierarchy: representations of the DIKW hierarchy", in: Journal of Information Science 33 (2007), 163 – 180.

Shannon, Claude Elwood (1948): "A Mathematical Theory of Communication", in: *Bell System Technical Journal*, 1948 (Juli, Oktober), 379–423, 623–656.

Thaller, Manfred (2009a): "The Cologne Information Model: Representing information Persistently", in: Thaller 2009b, 223 - 240.

Thaller, Manfred (2009b): *The eXtensible Characterisation Languages – XCL,* Hamburg.