

## Digital Humanities in Jena

**Veranstalter:** DHnet Jena, Friedrich-Schiller-Universität Jena

**Datum, Ort:** 23.11.2017, Jena

**Bericht von:** Martin Prell, Historisches Institut, Friedrich-Schiller-Universität Jena

Am 23. November 2017 veranstaltete das Netzwerk zur Förderung der Digital Humanities in Jena (DHnet Jena<sup>1</sup>) einen Thementag, der dem Austausch und der Präsentation verschiedener Digital-Humanities-Projekte der Universität Jena und darüber hinaus diente. ROBERT GRAMSCH (FSU Jena) stellte das DHnet, das sich als Forum zur Vernetzung von DH-Interessierten aus der Studierenden- bis hin zur Professorenschaft, den Archiven und Bibliotheken sowie weiteren inner- und außeruniversitären Einrichtungen fest etabliert hat, zunächst kurz vor, bevor er Herausforderungen, Chancen aber auch Gefahren beim Aufbau der Digital-Humanities-Forschung und -Lehre in Jena erläuterte. Dem DHnet als quer zu den institutionellen Strukturen und Fachdisziplinen agierendem Akteur komme hierbei eine zentrale Rolle zu, nicht zuletzt um mögliche Entwicklungen der Digital Humanities hin zu einem „wissenschaftlichen Ghetto einer neuen Arkandisziplin“ frühzeitig zu verhindern, so Gramsch.

Im ersten Vortrag des Tages gaben PETRA KUNZE und ANGELA HAMMER (Thüringer Universitäts- und Landesbibliothek Jena, ThULB) Einblicke in den derzeitigen Stand der Kulturgutdigitalisierung im Freistaat. Die ThULB nehme eine bedeutende Stellung innerhalb eines umfassenden Kooperationsnetzwerkes zahlreicher Thüringer Kulturgutdigitalisierungspartner ein, das sich zunehmend vergrößere und stetig neue Wege der Zusammenarbeit und Aufbereitung erprobe, um passgenaue Abstimmungen auf die Bedürfnisse und Möglichkeiten der Partner vornehmen zu können. Letztere können dafür ein sehr umfassendes Leistungsportfolio der ThULB (Digitalisierung, Restaurierung, erschließendes, fachliches und juristisches Know How, technische Infrastruktur) in Anspruch nehmen. Die Referentinnen stellten exemplarische Portale und Inhalte aus dem

über 8 Mio. Digitalisate umfassenden digitalen Bestand der ThULB vor. Präsentiert wurden unter anderem die stark frequentierten Plattformen Journals@UrMEL<sup>2</sup> oder Collections@UrMEL<sup>3</sup>, die zentrale Instanz für historische Drucke, Urkunden Handschriften und (teils unikale) Karten (auch kleinerer Einrichtungen). Diese bilden allerdings nur einen Bruchteil der über die Jahre entstandenen zahlreichen, auch institutionen- und projektspezifischen Fachportale (z.B. Digitales Archiv des Landesarchivs Thüringen<sup>4</sup>, Goethe- und Schiller-Archiv<sup>5</sup>, Digitale Historische Bibliothek Erfurt/Gotha<sup>6</sup> und viele digitale Editionen<sup>7</sup>), deren Forschungsdaten auch in übergreifenden digitalen Sammlungsportalen wie dem Archivportal-D, Kalliope, ZDB, ZVDD u.a. recherchierbar sind. Viele dieser Portale ermöglichen zudem erweiterte Zugänge zum historischen Kulturgut bspw. durch deren Visualisierung auf Karten oder Zeitleisten, die die ThULB laufend und projektspezifisch entwickelt. Neben zweidimensionalen Schriftquellen gehören längst auch audiovisuelle und museale Quellen und Objekte dazu. Nicht zuletzt aufgrund der Erschließungsleistungen, die die ThULB auch für externe Bestände anbietet, und ihres umfassenden Serviceangebots von der Planung beginnender bis hin zur Kuration abgeschlossener Projekte ist sie als ein zentraler Eckpfeiler der Kulturgutdigitalisierung in Thüringen und darüber hinaus nicht wegzudenken.

Über neue Möglichkeiten der Erschließung von Karten und Globen und deren Umsetzung in Jenaer Projekten referierte ANDREAS CHRISTOPH (FSU Jena). Ausgehend von den zahlreichen Informationen, die eine historische Karte beherberge und die es zu extrahieren gelte, problematisierte der Referent die Schwierigkeiten bei der Erhebung von Metadaten und betonte die Notwendigkeit von Mindestmetadatenstandards in die-

<sup>1</sup> <http://dhnet.uni-jena.de/> (Alle Urls zuletzt geprüft am 18.01.2018).

<sup>2</sup> <http://zs.thulb.uni-jena.de/content/below/index.xml>

<sup>3</sup> [https://archive.thulb.uni-jena.de/hisbest/templates/master/template\\_collections/index.xml](https://archive.thulb.uni-jena.de/hisbest/templates/master/template_collections/index.xml)

<sup>4</sup> <https://archive.thulb.uni-jena.de/staatsarchive/>

<sup>5</sup> <https://archive.thulb.uni-jena.de/gsa/>

<sup>6</sup> <https://archive.thulb.uni-jena.de/ufb/>

<sup>7</sup> <http://www.urmel-dl.de/#projects>

---

sem Bereich. Dies werde insbesondere dann offensichtlich, wenn es darum gehe, die Datenbestände einer Institution mit denen anderer Einrichtungen und deren Datenstandards zu verknüpfen. Christoph stellte daher ein einheitliches Modell zur Datenerfassung für Bibliotheken vor, das er derzeit erarbeitet. Ein wichtiger Verarbeitungsschritt zur (Nach)Nutzung von Karten in DH-Kontexten bildet die Georeferenzierung, deren Herausforderungen der Referent skizzierte. Die Lagegenauigkeit von Passpunkten, die Detailtreue des Kartenmaterials oder die vermessungstechnischen Grundlagen einer Karte bilden hierbei nur einige zentrale Faktoren, die sowohl die Georeferenzierung selbst als auch die anschließende Nutzbarkeit einer Karte deutlich beeinflussen. Abschließend gab Christoph Einblicke in einzelne Kartenprojekte, an denen er mitwirkt, wie bspw. das Portal „Kartenspeicher“<sup>8</sup>, das aufgrund zu vieler Inselfösungen notwendig wurde und sich nicht zuletzt durch eine den Gemeinsamen Bibliotheksverbund (GBV) übergreifende Erfassung auszeichnet, oder die Erstellung von 3D-Scans samt anschließender Viewer-Darstellung.

Waren die ersten Vorträge auf die Erhebung, Aufbereitung und Präsentation von Kulturgutdaten ausgerichtet, so fokussierte der Workshop mit dem Vortrag von PHILIPP POSCHMANN und JAN GOLDENSTEIN (beide FSU Jena) nun verstärkt auf forschungsfragengeleitete Analysen bestehender Daten. Das Ziel des Projektes der beiden Wirtschaftswissenschaftler ist es, Wikipedia zur Entwicklung betriebswirtschaftlicher Lösungen zu nutzen und dafür insbesondere handelnde soziale Akteure auffindig zu machen. Die Referenten stellten zentrale Werkzeuge vor, mit denen sie Personen in Wikipediaeinträgen identifizieren (Named-Entity-Recognition und Disambiguierungsverfahren) und benannten dabei auch Bedingungen und Grenzen dieser Verfahren. Einer allgemeinen Skepsis gegenüber Wikipedia als geeignete Datenbasis für wissenschaftliche Auswertungen begegneten Poschmann und Goldenstein mit Studienergebnissen, nach denen die durchschnittliche Qualität von Wikipedia-Artikeln im Vergleich zu alternativen Ressourcen gut sei, sodass weni-

ger fehlerhafte Artikel – die zudem nachbearbeitet werden könnten – als vielmehr lückenhafte Beiträge ein Problem darstellten. Am Beispiel der Berichterstattung der Haupt-US-Newspaper über die Präsidentschaftswahlen in den USA der Jahre 2015/16 demonstrierten die Referenten ihr Vorgehen, das aus dem Dreischritt Detecting, Disambiguating und Specifying Social Actors besteht. Angewendet auf die semantisch verknüpfende Wissensbasis DBPedia<sup>9</sup> und unter zu Hilfenahme des Tools „DBPedia Spotlight“ ließen sich so aus dem öffentlichen Wahlkampfdiskurs gezielt demografische Informationen zu Personen und Unternehmen gewinnen – mit einer Genauigkeit im 90er-Prozent-Bereich. Abschließend erläuterten die Referenten die Anschlussmöglichkeiten weiterer DH-Verfahren und strichen die einzelnen Vor- und Nachteile von Wikipedia als Datenbasis heraus.

Ebenfalls dem Bereich des Text-Mining zuzuordnen war der Vortrag von JOHANNES HELLRICH (FSU Jena) zur Veränderung von Wortbedeutungen im Laufe der Zeit. Ausgehend von dem Credo Ludwig Wittgensteins, die Bedeutung eines Wortes sei sein Gebrauch in der Sprache, betrachtet Hellrich in seinem Projekt das Wortumfeld gesuchter Begriffe und fokussiert dabei zunächst auf jene Wörter, die nicht und jene, die häufig miteinander vorkommen. Die Ähnlichkeit von Wörtern werde schließlich sichtbar durch das Aufzeigen eines ähnlichen Wortumfeldes. Beispielhaft führte der Referent die Messung diachroner Wortsemantik an den Begriffen „Gay“ und „Herz“ vor und visualisierte den Bedeutungswandel des Herz-Begriffes durch die Sichtbarmachung des Wandels ähnlicher Wörter, die sich vom Kontext „Gemüth“ weg und hin zu Begriffen wie „Gehirn“ oder „Lunge“ entwickelten. Mit JeSemE<sup>10</sup> hat Hellrich zudem einen Webdienst erstellt, mit dem sich auch ohne Programmierkenntnisse und auf der Grundlage fünf auswählbarer Textkorpora der Wandel der Wortbedeutung zahlreicher Wörter für die eigenen Forschungsfragen nachverfolgen lässt. NutzerInnen können hier zwischen zwei zum Einsatz kommenden Algorithmen des Word Embedding (Sin-

<sup>8</sup> <http://kartenspeicher.gbv.de/>

<sup>9</sup> <http://wiki.dbpedia.org/>

<sup>10</sup> <http://jeseme.org/>

gularwertszerlegung und word2vec) auswählen, deren Stärken und Schwächen Hellrich skizzierte.

Der linguistischen Annotation geschriebener Sprache widmete sich der Vortrag von BARBARA AEHNLICH und HENRY SEIDEL (beide FSU Jena). Den Ausgangspunkt ihrer Ausführungen bildete der Umstand, dass existierende Tagging-Tools in der Regel auf verschiedene Sprachstufen zugeschnitten seien, was durch Unterschiede in der Phonologie und Graphologie bedingt sei. Zudem seien die derzeitigen Tools häufig auf Sprachdaten der Gegenwartssprache trainiert, so dass deren Anwendung auf frühneuhochdeutsche Texte zu einer korrekten Lemmatisierungsrate von nur ca. 60 Prozent führe, so die Referenten. Mit der Software LAKomp<sup>11</sup> stellen Aehnlich und Seidel ein Tool zum halbautomatischen Taggen des Frühneuhochdeutschen vor. Für dessen Anwendung sei es zunächst erforderlich, die Texte manuell zu transkribieren, da die automatische Erkennung via OCR- oder HTR-Technologien für die beabsichtigten Zwecke ungenügende Ergebnisse liefere. Nachfolgend lemmatisierte und POS-tagge (nach STTS- oder HiTS-Konventionen) die Software halbautomatisch die Transkription, die bis auf die Auflösungen der Abkürzungen und Eingriffe in die Interpunktion keine weiteren Korrekturen erfordere. LAKomp charakterisiere zudem eine sehr einfache Bedienbarkeit und die Möglichkeit orts- und zeitunabhängigen Taggens. Mit EXMARALDA<sup>12</sup> stellten die Referenten zudem eine zweite, XML- und Java-basierte Software zur Annotation des Frühneuhochdeutschen vor und betrachteten diese in ihren Einzelteilen. Dabei strichen sie auch wichtige Unterschiede dieser zu weiterer etablierter Tagging-Software wie bspw. dem TreeTagger heraus, der erst für jüngere Sprachstufen geeignet sei und daher eine vorherige intensive Normalisierung der zu untersuchenden Texte erfordere. Den Abschluss des Vortrages bildete ein Vergleich zwischen LAKomp und EXMARALDA hinsichtlich deren jeweiligen Stärken und Schwächen in der Anwendung.

SVEN BÜCHEL (FSU Jena) referierte zu einem Thema aus dem Forschungsbereich der Sentiment Analysis. Das Ziel seines Projektes ist es, Emotionen in historischer Spra-

che, zu denen bisher nur sehr wenige Daten vorhanden seien, mit quantifizierenden Methoden zu bestimmen. Büchel erläuterte zunächst das methodische Vorgehen seiner Arbeit, in der er auf das VAD-Modell zur formalen Darstellung von Emotionen zurückgreift. Dabei wird die Emotionalität eines Begriffes mit Hilfe der Parameter Valence, Arousal und Dominance ermittelt. Emotionslexika, in diesem Falle das Deutsche Emotionenlexikon ANGST, liefern als Grundlage eine zeitgenössische emotionale Bewertung zahlreicher deutscher Wörter. Um dieses Lexikon auch auf historische Sprache anwenden zu können, müsse es erweitert werden. Dazu setzt Büchel Verfahren des Word-Embedding ein, um über die Bestimmung ähnlicher Wörter deren (ggf. variierende) Emotionalität über die Zeit hinweg zu bestimmen. Die Quantifizierung wiederum erfolge mit Hilfe des Turney-Littman-Algorithmus'. Büchel konnte so sieben verschiedene Emotionslexika der Jahre 1690 bis 1899 mit Zeiträumen von je 30 Jahren erstellen. Die theoretischen Ausführungen illustrierte der Referent anschließend am konkreten Beispiel der emotionalen Varianz des Begriffes Sünde über 200 Jahre hinweg auf Grundlage des Textkorpus' des Deutschen Textarchivs. Mit dem von ihm entwickelten kommandozeilenbasierten Tool JEmAS<sup>13</sup> präsentierte Büchel zudem einen nachnutzbaren lexikonbasierten Ansatz zur Emotionenmessung. Zum Abschluss des Vortrages demonstrierte der Referent am Beispiel der Entwicklung literarischer Gattungen und Textsorten deren emotionalen Zuschreibungswandel vom Ende des 17. bis Ende des 19. Jahrhunderts.

Als Referent für den Abendvortrag und Abschluss des Thementages konnte TORSTEN SCHRADE (AdW Mainz sowie FH Mainz) gewonnen werden, der zum Thema Linked Open Data sprach. Den Ausgangspunkt seiner Ausführungen bildeten Problematisierungen der Vielzahl der verschiedenen existierenden Datenquellen und -formate und die heterogenen Modellierungsweisen gleicher Phänomene – beispielweise in der TEI – mit dem Ergebnis häufig nur unzureichend oder gar

<sup>11</sup> <http://lakomp.uzi.uni-halle.de/>

<sup>12</sup> <http://exmaralda.org/de/>

<sup>13</sup> <https://github.com/JULIELab/JEmAS>

---

nicht miteinander vernetzter Forschungsdaten. Diesem Befund stellte Schrade zunächst einführende Betrachtungen zu historischen Klassifikationsansätzen gegenüber, wie sie bei antiken und mittelalterlichen Gelehrten anzutreffen seien, um die historische Tradition der Prinzipien von Linked-Open-Data aufzuzeigen. Mit der Erläuterung des „Five-Star“-Prinzips, den fünf LOD-Empfehlungen Tim Berner-Lees für die Erzeugung von Linked Open Data, leitete Schrade schließlich zu den Grundlagen des Semantic Web über. Beispielhaft und detailliert demonstrierte der Referent das Funktionsprinzip von Triplen, bevor er auf deren Verwendung im Rahmen des RDF-Datenmodells und die Notwendigkeit von Vokabularen und Ontologien für die maschinelle Auswertbarkeit der Triple zu sprechen kam. Dieses illustrierte er schließlich an einem Brief J. W. v. Goethes an F. W. Riemer, das im CMI-Datenformat (einer Mini-Ontologie für Briefkorrespondenzen, wie sie bspw. im Webservice *correspSearch*<sup>14</sup> zum Einsatz kommt) erfasst und durch die Triple-Kodierung die Generierung zusätzlicher inhärenter Erkenntnisse ermöglichte. Nach diesen grundlegenden Ausführungen zur Thematik gab Schrade Einblicke in die Einsatzmöglichkeiten von LOD und Semantic Web in verschiedenen Projekten. Dazu gehörten beispielsweise die *Epidat*-Forschungsplattform für jüdische Grabsteingraphik<sup>15</sup> unter Einsatz der Erlangen-CRM-Ontologie oder der *Generic Viewer*<sup>16</sup>, ein Tool zur semantischen Anreicherung von 3D-Daten. Den Abschluss des Vortrages bildeten allgemeine Reflexionen zu den Potentialen und Vorteilen des Einsatzes von LOD für die Analyse von Forschungsdaten, deren Vernetzung und Nachnutzung.

Der Workshop, der neben den Vorträgen auch die Gelegenheit zum Kennenlernen weiterer Thüringer DH-Projekte und ihrer Akteurinnen und Akteure in Form von Posterpräsentationen bot<sup>17</sup>, hat eindrücklich die Vielfältigkeit nicht nur der Jenaer, sondern der Thüringer Digital-Humanities-Landschaft allgemein demonstriert. Das zahlreiche und sehr heterogene Publikum und die sehr angeregten Diskussionsrunden zeugten von dem hohen Interesse und der Relevanz, die computergestützte Forschungsmethoden im Freistaat genießen. Das DHnet Je-

na hat hierbei erneut seine Bedeutung als ein wichtiger Eckpfeiler für den interdisziplinären Austausch von Forschungsvorhaben und -ergebnissen aus dem Bereich Thüringer Digital-Humanities-Forschung demonstriert. Bleibt zu wünschen, dass der Workshop zu weiteren Kooperationen und Projektideen angeregt hat und er zukünftig Fortsetzungen erfährt.

#### Konferenzübersicht:

Petra Kunze / Angela Hammer (Jena): Zum Stand der Kulturgutdigitalisierung in Thüringen

Andreas Christoph (Jena): Perspektiven auf Karten und Globen | Virtuelle Weltenmodelle in 2D und 3D

Philipp Poschmann / Jan Goldenstein (Jena): Wikipedia als Quelle für demographische Informationen

Johannes Hellrich (Jena): Automatische Analyse (diachroner) Wortsemantik

Barbara Aehnlich / Henry Seidel (Jena): Händisches und halbautomatisches Taggen

Sven Büchel (Jena): Quantifizierung von Emotionen in historischer Sprache

Torsten Schrade (Mainz): "Set your data free" - Vom geisteswissenschaftlichen Datensilo zur vernetzten Wissensbasis mittels Linked Open Data

Tagungsbericht *Digital Humanities in Jena*. 23.11.2017, Jena, in: H-Soz-Kult 27.02.2018.

---

<sup>14</sup> <http://correspsearch.net/>

<sup>15</sup> <http://www.steinheim-institut.de/cgi-bin/epidat>

<sup>16</sup> <http://www.spatialhumanities.de/ibr/technologie/genericviewer.html>

<sup>17</sup> Die Vorträge und Poster des Thementages sind unter folgender URL einsehbar: [https://www.db-thueringen.de/receive/dbt\\_mods\\_00033310](https://www.db-thueringen.de/receive/dbt_mods_00033310)