

Workshop eComparatio: Textvergleich und digitaler Apparat

Schubert, Charlotte

schubert@uni-leipzig.de

Universität Leipzig\Historisches Seminar, Deutschland

Kahl, Hannes

hannes.kahl@uni-leipzig.de

Universität Leipzig\Historisches Seminar, Deutschland

Meins, Friedrich

friedrich_meins@uni-leipzig.de

Universität Leipzig\Historisches Seminar, Deutschland

Bräckel, Oliver

oliver.braeckel@uni-leipzig.de

Universität Leipzig\Historisches Seminar, Deutschland

I. Motivation

In den textbasierenden Geisteswissenschaften ist eine optimale Beschaffenheit der zu Grunde liegenden Texte eine wesentliche Bedingung für wissenschaftliches Arbeiten. Wie sehr sich die Bedeutung eines Textes schon durch scheinbar minimale Unterschiede wie die der Interpunktion verschieben kann, hat sich einer breiteren Öffentlichkeit zuletzt in der Diskussion über einen Punkt in einer Abschrift der amerikanischen Unabhängigkeitserklärung gezeigt (The Atlantic 7/2014).

Während die historische Diskussion über digitale Editionen vielfach mit dokumentarischen Editionstypen bzw. der Frage nach dem Dokumentcharakter der Grundlagen einer Edition (Manuskripte etc.) und der Frage nach der Essenz des Textbegriffes beschäftigt ist (P. Sahle 2013), ergibt sich insbesondere für diejenigen historischen und alttumswissenschaftlichen Disziplinen, die weiterhin auf die klassische Form kritischer Texteditionen angewiesen sind, ein anderes Problem: Da sie traditionell in hohem Maße mit Texten beschäftigt sind, die notwendig als Interpretationen und Rekonstruktionen gelten müssen (West 1973,32), ist hier vor allem die Frage nach der Beschaffenheit und Begründung der zum Teil massiven editorischen Eingriffe nicht nur in die Lesart, sondern auch in den Umfang und die Zuschreibung von Texten, etwa im Falle sogenannter „Fragmente“, von vorrangiger Bedeutung. Der Workshop soll in diese allgemeine Problematik einführen und sich im Hinblick auf das Thema der Konferenz dem Bereich Kritik der digitalen Geisteswissenschaften (traditionelle

Fächer und DH) zuordnen. Den Teilnehmenden soll dies praktisch anhand der Anwendung der Software eCOMPARATIO vermittelt werden. eCOMPARATIO bietet eine einfache Möglichkeit der Kollationierung verschiedener Varianten eines Textes und ermöglicht eine digitale, auf dem automatischen Textvergleich beruhende Form des kritischen Apparates.

Dazu sind einzelne Use Cases ausgearbeitet worden (anhand der Fragmentsammlung der Vorsokratiker von Diels/Kranz [griechisch], der Res Gestae des Augustus [lateinisch], des Genfer Gelöbnisses [deutsch], der Gettysburg Address von Abraham Lincoln [englisch]), anhand derer den Teilnehmern die Funktionalitäten demonstriert werden.

II. Die Software

Im Projekt eCOMPARATIO, das von 2014 bis 2016 von der DFG gefördert wurde und in Leipzig am Lehrstuhl für Alte Geschichte in Kooperation mit dem Center of E-Humanities in History and Social Sciences (ICE) am Max-Weber-Kolleg für kultur- und sozialwissenschaftliche Studien durchgeführt wurde, ist vor dem Hintergrund dieser Problematik ein einfach zu bedienendes Tool für den Vergleich prinzipiell beliebig vieler und beliebig langer digitalisierter Editionen vorrangig griechischer und lateinischer, technisch gesehen aber auch sämtlicher anderer in einem UNICODE-Format zugänglicher Texte entstanden. Der Textvergleich arbeitet auf der Basis der Identifikation von Ungleichheiten. Hierbei reicht die Spanne der programmiertechnisch unterscheidbaren Differenzen von Ungleichheiten innerhalb von Wörtern und Buchstabenfolgen bis hin zu vertauschten Passagen. Im Unterschied zu anderen Vergleichsprogrammen benötigen Anwender weder eine Installation von Python, eine Levenshtein Bibliothek oder ein Java-Plugin, sondern erhalten eine für die Anwender ohne Vorkenntnisse sofort im Browser nutzbare und mit Copy-and-Paste einfach zu bedienende Oberfläche. Auch Fallbeispiele, eine Textdokumentation sowie Videoanleitungen stehen zur Verfügung. In zwei derzeit (2016-17) von der DFG und der Andrew W. Mellon Foundation geförderten Projekten an der Universität Leipzig in Kooperation mit Christopher Blackwell von der Furman University in Greenville, SC/ USA, erfolgt eine Einbindung des Vergleichstools in ein Interface (zu dem Protokoll Canonical Text Services (CTS) als Teil der CITE Architecture: <http://cite-architecture.github.io/cts/>), dessen Ziel nicht nur die Bereitstellung möglichst vieler digitalisierter Editionen einzelner Texte, sondern damit auch erweiterte Möglichkeiten des Textvergleiches, der Suche nach Parallelstellen und weiterer Formen des sog. „Text-Mining“ ist.

Die Texteingabemaske

Die eigenständige Version der Vergleichssoftware (unter <http://ecomparatio.net/~khk/instanzen/ecompp/> ist ein Beispiel frei zum Gebrauch bereitgestellt) ermöglicht weiterhin eine browserbasierte oder auch offline verwendbare Anwendung für Texte nach dem Copy-and-Paste-Prinzip: Hier können beliebig viele Versionen eines Textes eingefügt werden.

Die Darstellung

Das Tool ermöglicht verschiedene Ausgaben des Vergleiches:

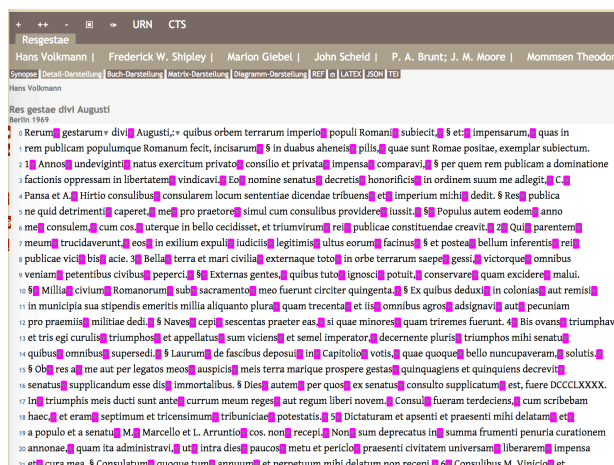


Abbildung 1, Detail-Vergleich

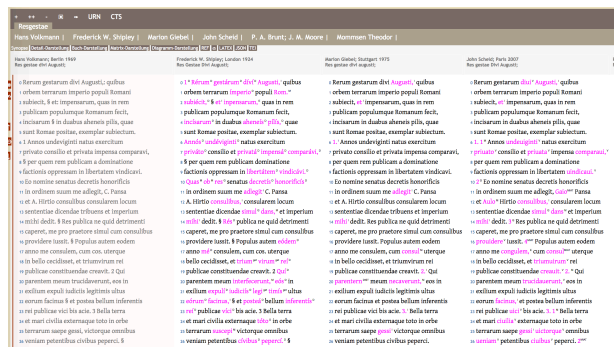


Abbildung 2, Synopsis

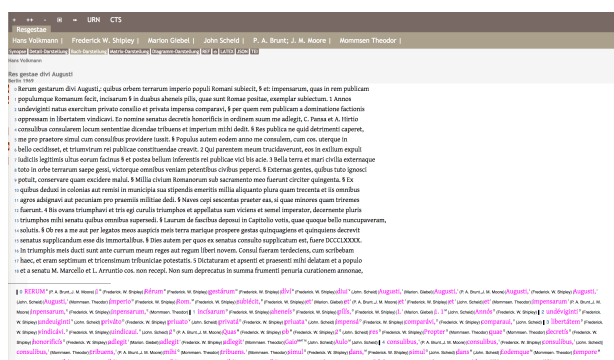


Abbildung 3, Buch-Darstellung mit digitalem Apparat zum Textvergleich

Ausgabe/Export der Ergebnisse

Die Ausgabe der Ergebnisse kann zur weiteren Integration in den Arbeitsprozess als TEI XML oder LaTeX Code erfolgen. Will man die Vergleichsdaten abfragen, so steht ein JSON Interface zur Verfügung, über das weitere Software angebunden werden kann.

III. Ziele und Zielgruppe

Ziel des Workshops ist eine Einführung in die Anwendung der eCOMPARATIO Vergleichssoftware, auch in Abgrenzung zu bereits verfügbarer Software (collateX, Juxta) mit ähnlichen Anwendungsbereichen. Dazu soll zunächst eine Einführung in die Relevanz der Frage nach scheinbar marginalen textlichen Unterschieden anhand prägnanter Beispiele aus verschiedenen historischen Epochen und in verschiedenen wissenschaftlichen Diskurssprachen gegeben werden (anhand griechischer, lateinischer, englischer und deutscher Texte).

Darüber hinaus soll den Teilnehmenden die Möglichkeit gegeben werden, auch Texte aus den eigenen Disziplinen oder beliebige im Internet verfügbare Versionen von Texten selbst miteinander zu vergleichen und sich so mit den Funktionen der Software vertraut zu machen.

Zielgruppe des Workshops sind prinzipiell alle Interessierten aus dem Bereich der textbasierten Geisteswissenschaften. Dabei sind diejenigen, die, wie oben angesprochen, vor allem mit der Vielzahl digitalisierter Editionen arbeiten, ebenso angesprochen wie solche, die selbst an der Erstellung von Editionen etc. arbeiten, und für die eCOMPARATIO ein hilfreiches Mittel bei der Sichtung und Kollationierung von Textzeugnissen sein kann.

Besondere technische Kenntnisse sind nicht erforderlich, da sich eCOMPARATIO bewusst an Anwender richtet, die entweder selbst an einer kritischen Edition arbeiten oder auf der Grundlage mehrerer kritischer Editionen wissenschaftliche Fragestellungen verfolgen.

IV. Ablauf und Teilnehmerzahl

Im Workshop soll die browsergestützte Version zur Anwendung kommen, ein Download der Software wird aber auch möglich sein. Neben der bereits beschriebenen Einführung in die wissenschaftliche Grundproblematik erfolgt zunächst eine kurze Präsentation eigener Ergebnisse im Zuge der Forschung mit eCOMPARATIO.

Hauptsächlich soll im praktischen Teil den Teilnehmenden die Möglichkeit gegeben werden, auf ihren

eigenen Rechnern in der browsergestützten Version selbst Vergleiche der für sie relevanten Texte vorzunehmen, die mitgebracht werden sollten oder vor Ort aus Onlinedatenbanken heruntergeladen werden können.

Während des Workshops soll dann von Seiten der Organisatoren auf eventuelle individuelle Probleme und Schwierigkeiten eingegangen werden. Die Organisatoren des Workshops erhoffen sich hiervon einen eigenen Erkenntnisgewinn auch im Hinblick auf die Anwendungsmöglichkeiten und notwendige Ergänzungen im Hinblick auf die Arbeit gerade mit nicht-lateinischen oder nicht-griechischen Texten.

V. Ablauf

Die TeilnehmerInnen erhalten vorab eine detaillierte Anleitung (Handbuch eCOMPARATIO).

Im eigentlichen Workshop werden die jeweiligen Arbeitsschritte von einem der Organisatoren live vorgeführt (dafür wird ein leistungsstarker Beamer benötigt). Die konkreten Inhalte orientieren sich dabei an den bisher von den Organisatoren ausgearbeiteten Use Cases (s.o.), die von den Organisatoren präsentiert werden.

Während des Workshops werden wir bei auftretenden Fragen und Problemen den Teilnehmenden helfend zur Seite stehen, da sie auch die Möglichkeit haben sollen, anhand eigener Texte zu arbeiten. Um eine möglichst gute Betreuung der TeilnehmerInnen gewährleisten zu können, sollte die Teilnehmerzahl 25-30 nicht überschreiten.

VI. Organisatoren

Charlotte Schubert ist Althistorikerin, hat zu Themen der Mentalitätsgeschichte, Medizin- und Wissenschaftsgeschichte sowie zu verschiedenen Bereichen der griechischen Geschichte gearbeitet; seit 2006 verantwortliche Koordinatorin in verschiedenen DH-Projekten, die vom BMBF (eAQUA, eXChange), der DFG (eCOMPARATIO, CTS, Etablierung eines Open Access Online eJournals: Digital Classics Online), der VolkswagenStiftung (Digital Plato) und der Andrew W. Mellon Foundation (CTS) gefördert wurden und werden.

Hannes Kahl ist Informatiker, Berufserfahrung aus diversen DH-Projekten, Entwickler von eCOMPARATIO, arbeitet an der Weiterentwicklung von CTS und an einer Dissertation zu dem Thema „Form und Formalisierung – mit Anwendung innerhalb automatischer Ermittlung von Buchstabenwerten aus digitalen Abbildungen griechischer Lettern innerhalb wissenschaftlicher Editionen sowie deren digitaler Formatierung“ (Betreuer: Prof. Ch. Schubert, Alte Geschichte/Universität Leipzig/ Prof. O. Arnold, Informatik/FH Erfurt).

Friedrich Meins ist Althistoriker und hat eine Dissertation zum Thema „Literarische Kritik, rhetorische Theorie und historische Methode bei Dionysios von Halikarnassos“ geschrieben. Im Bereich der eHumanities hat er in den

Projekten „eAQUA“ und „eAQUA Dissemination“ an der Uni Leipzig sowie im Projekt „eCOMPARATIO“ am ICE der Universität Erfurt mitgearbeitet. Derzeit arbeitet er im von der Andrew W. Mellon Foundation geförderten Kooperationsprojekt der Universität Leipzig und der Furman University (Greenville, SC/ USA) „Annotating and Editing With Canonical Text Services (CTS)“.

Oliver Bräckel ist Althistoriker und arbeitet an einer Dissertation zu Politischen Flüchtlingen im Römischen Reich. Im Bereich der eHumanities hat er im Projekt „eCOMPARATIO“ am ICE der Universität Erfurt sowie im Projekt eXChange an der Uni Leipzig mitgearbeitet. Derzeit arbeitet er im von der Andrew W. Mellon Foundation geförderten Kooperationsprojekt der Universität Leipzig und der Furman University (Greenville, SC/ USA) „Annotating and Editing With Canonical Text Services (CTS)“.

Bibliographie

A. Olheiser, Have We Been Reading the Declaration of Independence All Wrong?, <https://www.theatlantic.com/entertainment/archive/2014/07/typo-could-mean-weve-been-reading-the-declaration-of-independence-all-wrong/373915/> (letzter Zugang 7.7.2017).

P. Sahle, Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels, Norderstedt 2013, 3 Bde. (Schriften des Instituts für Dokumentologie und Editorik Bd. 7).

M.L. West, Textual Criticism and Editorial Technique, Stuttgart 1973.

Links:

The Atlantic: <https://www.theatlantic.com/entertainment/archive/2014/07/typo-could-mean-weve-been-reading-the-declaration-of-independence-all-wrong/373915/> (letzter Zugang 7.7.2017)

<http://www.eaqua.net>

<http://www.ecomparatio.net/>

<http://www.ecomparatio.net/~khk/instanzen/ecompp/>

(letzter Zugang 17.9.2017)

<http://digital-plato.org>

<http://digital-classics-online.eu>

<http://cite-architecture.github.io/cts/>