

Squirrel

Ein innovativer Editor zur Erstellung digitaler Editionen und sein Pilotprojekt¹

Mark Hengerer, Isabella Hödl-Notter, Gerhard Schön

Die Erstellung von Transkriptionen und die Veröffentlichung von Editionen gehören zum Kernbereich der Geschichtswissenschaft. Beide Tätigkeiten sind oftmals schwierige und langwierige Prozesse und brauchen in der Lehre viel Zeit, individuelle Übung und professionelle Unterstützung.

Die kritische Edition handschriftlicher Quellen, einst von nicht wenigen als altmodisch und wenig kreativ belächelt, erfährt durch die neuen Möglichkeiten von Open-Access-Publikationen seit einigen Jahren eine neue, sich in zahlreichen digitalen Editionen äußernde Akzeptanz. Diese Entwicklung geht allerdings noch nicht überall mit einer Stärkung der grundwissenschaftlichen Ausbildung an den deutschsprachigen Universitäten einher; erst 2015 in einem Grundsatzpapier des Verbandes der Historiker und Historikerinnen Deutschlands wurde beklagt, dass entsprechende Ausbildungsangebote zum Umgang mit handschriftlichen Quellen rückläufig seien.²

-
- 1 Für konstruktive Anregungen zu diesem Text und die großzügige Unterstützung des Projektes danken wir Dr. Christian Riepl (IT-Gruppe Geisteswissenschaften).
 - 2 „Grundlegende Kenntnisse und Fähigkeiten drohen deshalb nicht nur bei den Studierenden, sondern langfristig auch bei den Lehrenden in einem Maße abzunehmen, dass die kulturelle Überlieferung der Vergangenheit nicht mehr eigenständig erschlossen und beurteilt werden kann.“ Forum: E. Schlotheuber / F. Bösch: Quellenkritik im digitalen Zeitalter: Die Historischen Grundwissenschaften als zentrale Kompetenz der Geschichtswissenschaft und benachbarter Fächer, in: H-Soz-Kult, 16.11.2015, <http://www.hsozkult.de/debate/id/diskussionen-2866> (30. Oktober 2017).

Um die geschichtswissenschaftliche Kernkompetenz im Bereich Transkription und Edition zu stärken, startete im Jahr 2014 am Historischen Seminar der LMU München ein Editionsprojekt auch als informatisches, von der IT-Gruppe Geisteswissenschaften betreutes Pilotprojekt.³ Der Editor, den wir mit Blick auf die Freuden und Leiden des wissenschaftlichen Edierens inzwischen *Squirrel* genannt haben, wurde im Wintersemester 2014/15 von Gerhard Schön (IT-Gruppe Geisteswissenschaften) in Zusammenarbeit mit den Studierenden des Editionsprojekts „Die Briefe Erzherzog Leopold Wilhelms an Kaiser Ferdinand III. aus dem Reichsarchiv Stockholm“ – Isabella Hödl-Notter, Franziska Honer, Julius Kiendl, Heike Nickel, Hildegard Renner, Alexandra Röckel, Christine Rogler und Pia Zachary unter Leitung von Mark Hengerer – entwickelt. Seit Frühjahr 2017 besteht zudem eine Kooperation mit dem Computerlinguisten Max Hadersbeck und seinem Team, Oksana Budurova, Azada Rustamova, Stefan Schwyter und Abuliezi Xieyidan. Ziel dieser vom Programm Lehre@LMU finanziell geförderten Kooperation ist die Verschränkung des Editors mit computerlinguistischen Analysetools (Unitex/GramLab). *Squirrel* trägt mithin dazu bei, zukunftssträchtige grundwissenschaftliche Forschungsstrukturen zu entwickeln und nachhaltig zu verankern.

Die IT-Gruppe Geisteswissenschaften wiederum bereitet auf einem fach- und fakultätsübergreifenden Konzept die Grundlagen für eine interdisziplinäre Zusammenarbeit von Geschichtswissenschaft, Informatik und Computerlinguistik, indem sie (a) dauerhaft die informationstechnische Infrastruktur (File-, Datenbank- und Webserver) zur Verfügung stellt, (b) Softwaremodule projektübergreifend konzipiert, entwickelt, wartet und pflegt sowie projektspezifisch adaptiert, (c) in Zusammenarbeit mit der Universitätsbibliothek und dem Leibniz-Rechenzentrum (LRZ) Einzelprojekte in ein umfassendes und nachhaltiges Forschungsdatenmanagement einbindet.

3 http://www.fnz.geschichte.uni-muenchen.de/forschung/forsch_projekte/briefe-leopold/index.html

1. Squirrel. Der Editor

1.1 Technische Details

Die Zweispaltenansicht stellt dem zoomfähigen Scan der Briefseite ihre Transkription gegenüber, die als Freitext erfasst und nach einem vorgegebenen, aber adaptierbaren Schema annotiert werden kann (Abb. 1). Die Wörter aus dem Faksimile können per Textrahmen mit ihrem Transkript verknüpft werden. Zu jedem markierten Wort werden dann automatisch die Bilder von gleich und ähnlich transkribierten Schriftzügen als Hilfestellung eingeblendet (Abb. 2). Der Editor *Squirrel* stellt so beim Edieren visuelles Vergleichsmaterial für transkribierte Passagen zur Verfügung und ermöglicht ein permanentes Feedback während der Transkriptionsarbeit. Die Orientierung an Vergleichsmaterial kann zwar zeitweise auch Falschlesungen stabilisieren, doch hat die technisch stark unterstützte Arbeit mit visuellem Vergleichsmaterial solche Fehler bislang regelmäßig auch aufzudecken geholfen.

Alle diese Bildausschnitte sind auch über eine Suchfunktion erreichbar und jeweils mit einem Link zur Sofortkorrektur der Transkription versehen (Abb. 3).



Abbildung 1



Abbildung 2

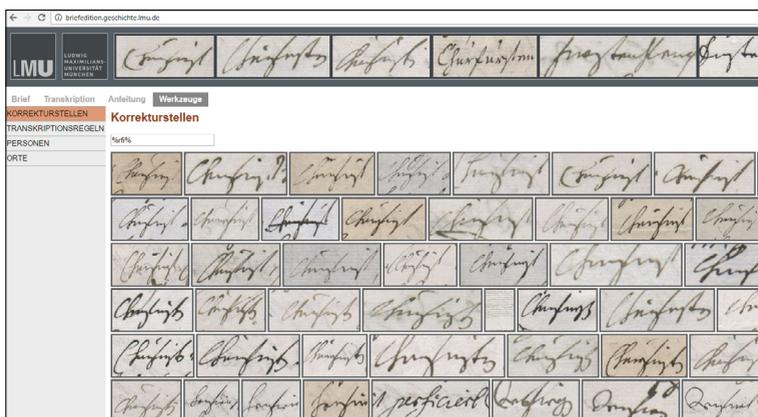


Abbildung 3

Squirrel basiert clientseitig vollständig auf HTML5 und erfordert auf Benutzerseite außer einem Browser und Internetverbindung keinerlei Zusatzsoftware oder Plugins. Serverseitig ist die Anwendung in PHP implementiert. Der Editionstext samt XML Annotationen wird auf dem Server vorgehalten (derzeit in geschredderter Form im relationalen Datenbanksystem MySQL) und kann über die Webseite als XML exportiert und durch XSLT in verschiedene Ausgabeformate transfor-

miert werden. Für die Vergleichsfunktionen hat sich der Einsatz von Trigrammanalysen bewährt.

Die Weiterentwicklung des technischen Bereichs von *Squirrel* verteilt sich seit dem Frühjahr 2017 auf drei Bereiche:

1. Zunächst implementieren wir die wichtigsten Teile des vom Deutschen Textarchiv (DTA) verwendeten Basisformats (DTABf), einer auch von der DFG und CLARIN-D empfohlenen Einschränkung der TEI Richtlinie P5 für historische Drucke. Hierzu werden bewährte Standards innerhalb der Geschichtswissenschaft (für die Epochen Mittelalter und Neuzeit) ermittelt und im Editor als Annotationsoptionen bereitgestellt. Die domänenspezifische Adaption aufwändiger Einzelprojekte lässt sich damit auf ein geringes Maß an Programmierarbeit reduzieren. Projekte innerhalb der klassischen Quellengruppen sollten mit dieser Vorarbeit ohne weitere oder wenigstens ohne aufwändige Adaption auskommen.

2. Zudem verschränken wir *Squirrel* mit dem UNITEX/GramLab (<http://unitexgramlab.org/>). Mit Elementen des GramLab wird eines der zentralen Probleme von Online-Publikationen gelöst, nämlich die automatisierte Bereitstellung eines Registers. Baut man (mit vertretbarem Arbeitsaufwand) ein Lexikon mit UNITEX/Gramlab auf, zeigt die Suche nach Grundformen (Lemmata) sämtliche Flektionsformen in der Trefferliste an. Zudem lässt sich eine Wortkonkordanz einfach erstellen.

3. Schließlich bereiten wir Schnittstellen zu wichtigen Publikationsformaten (*XHTML*, *LaTeX/PDF*) vor, welche hybride Veröffentlichungen erleichtern. Erforderlich ist die elektronische und typographische Adaption der durch unsere Auswahl aus dem DTA Basisformat sowie der in den verschiedenen Epochendisziplinen verwendeten weiteren Annotationstypen (Formatierungen) in für die Studierenden bereitzustellenden Annotationsoptionen (Formatvorlagen). Die besten Ergebnisse im Printbereich liefern derzeit Dateien in LaTeX/PDF, im Onlinebereich Dateien mit XHTML.

1.2 Workflow

Für die Arbeit mit *Squirrel* bietet sich somit der folgende Workflow an:

- Einrichtung eines Zugangs zum *Squirrel* Editor für Lehrende und Studierende, gegebenenfalls durch Anmeldung im Digital Humanities Virtual Laboratory (DHVLab, <https://dhvlab.gwi.uni-muenchen.de>).
- Anlegen eines Datensatzes pro Archivalieneinheit (Brief, Faszikel) und Eintragung der Metadaten zur Quelle.
- Hochladen der zu transkribierenden Digitalisate (zum Beispiel Scans der Handschriften im JPEG Format).
- Übungen im Transkribieren (optimalerweise Blended Learning), von *Squirrel* unterstützt durch Einblendung gleich und ähnlich transkribierter Buchstabenfolgen in der Oberleiste der Anwendung (Abb. 2). Es ist so der visuelle Abgleich möglich, welcher durch Nutzung der Funktion „Werkzeug“ intensiviert werden kann; das Programm erlaubt die Anzeige aller gleich oder ähnlich transkribierten Buchstabenketten (Abb. 3). Schreibvarianten werden dadurch gut sichtbar.
- *Optional* Fortentwicklung der Transkriptionen zur publizierbaren Edition, das heißt Fixierung der im Hinblick auf Rezipienten, Ressourcen und wissenschaftliche Adäquanz hin spezifizierten Grundsätze der Texterstellung; diese „Domänenspezifikation“ ist zwar nötig, wird aber durch das DTA Basisformat soweit möglich minimiert, um eine Anschlussnutzung zu erleichtern. Obschon Interoperabilität angestrebt wird, ist doch festzuhalten, dass absolut interoperable Datenbanken im Bereich Edition historischer Quellentexte bislang noch nicht vorliegen.
- Erstellung von Register und Konkordanz (Disambiguierung, Lexikon) sowie gegebenenfalls weiterer linguistischer Analysen mit UNITEX/GramLab.
- Publikation im Format XHTML und PDF sowie im Druck.⁴

⁴ Warum Druckfassungen nach wie vor wichtig sind, macht die 2017 auf politischen Druck Chinas hin erfolgte Löschung von Artikeln aus der Online-Zeitschrift *China Quarterly* deutlich: siehe *New York Times*, 18. August 2017 „Cambridge University Press Removes Academic Articles on Chinese Site“. Die Kehrtwende des Verlages

2. Lehre. Das Pilotprojekt

2.1 Entwicklungsgeschichte

Die Entwicklung des Online-Editors *Squirrel* baut auf mehrjährigen Vorarbeiten auf. Am Anfang stand auf eine von Mark Hengerer am Historischen Seminar der LMU München durchgeführte Lehrveranstaltung zu frühneuzeitlichen Quellen im Wintersemester 2014/15. Als sich abzeichnete, dass die Motivation für die Bearbeitung eines längerfristigen Pilot- und Publikationsprojekts gegeben war, begann die Entwicklung des Online-Editors durch Gerhard Schön unter laufender Berücksichtigung der sich aus der Praxis ergebenden Wünsche und Anregungen der Teilnehmer.

Inzwischen wurde der Editor in der Lehre des Historischen Seminars an der LMU mehrmals eingesetzt, auch in einem mediävistischen Pilotprojekt zur Urkundenedition (Julian Schulz, Magdalena Weiler). Im Wintersemester 2017/18 bieten Max Hadersbeck, Mark Hengerer und Gerhard Schön gemeinsam eine Lehrveranstaltung an, in der die computerlinguistische Erweiterung von *Squirrel* an einem weiteren Editionsfallbeispiel erprobt wird („Textedition und Textanalyse digital – eine Einführung“).

Der Editor wurde inzwischen auch mehrfach öffentlich präsentiert und diskutiert, unter anderem beim Historikertag Hamburg 2016 (Postersession).

Zuletzt wurde der Editor mitsamt dem Lehrprojekt ausführlich bei einem Workshop des dh muc (<https://dhmuc.hypotheses.org/>) im November 2016 präsentiert.⁵

macht nur noch sichtbar, wie manipulierbar der Zugriff auf wissenschaftliche Publikationen ist; siehe The Guardian, 22. August 2017 „Cambridge University censorship U-turn is censored by China“.

5 <https://dhmuc.hypotheses.org/workshop-digitale-editionen-und-auszeichnungssprachen>

2.2 Lehre im Wandel

Neben den technischen Neuerungen bringt der Einsatz digitaler Methoden tiefgreifende Änderungen für die Lehre mit sich. Das notwendigerweise interdisziplinär strukturierte digitale Arbeitsumfeld erfordert nicht nur Weiterbildung, sondern verlangt auch ‚neue‘ soziale Fähigkeiten. Diese Entwicklung ist insofern neu, da die Geschichtswissenschaften trotz verstärkter Projektforschung lange Zeit vorrangig auf Einzelarbeit ausgerichtet waren.

Durch den Einzug digitaler Methoden in die Lehre erweist sich das Erlernen dieser sozialen Fähigkeiten als Herausforderung, denn Team- und echte Projektarbeit gewinnen für den Erfolg von Forschung und Lehre an fundamentaler Bedeutung. Die für Arbeit in Teams und Projekten nötigen Kompetenzen aber sind bislang nicht Teil des Kanons universitärer Lehre.⁶ Anhand der verschiedenen oben genannten Pilotprojekte wurden die Herausforderungen und Potentiale sichtbar, die in den Bereichen Teamarbeit und Projektmanagement auf die Lehre zukommen.

Unsere Erfahrungen lassen sich ungefähr wie folgt fassen: Bedingung erfolgreicher Arbeit in Teams ist bewusst strukturierte Kommunikation. Diese muss den Spagat zwischen den Erfordernissen wissenschaftlicher Präzision, individueller Motivation und Effekten von Gruppenbildung bewältigen. Bewährt haben sich in dieser Situation

- die explizite Reflexion der verschiedenen Rollen und Motivationslagen der Teammitglieder – Zielkonflikte ließen sich so auf akzeptierte Ausgangsbedingungen zurückführen und konstruktiv lösen;
- strukturierte Interaktion, welche wegen der zeitintensiven dislozierten Arbeit im Browser nicht mehr selbstverständlich gegeben ist

⁶ Entsprechend attraktiv sind externe Kursangebote, hier am Beispiel der LMU: Graduate Center (ab Promotion/Promotionsplanung), LMU Center for Leadership and People Management (ab Promotion), LMU Extra (unbeschränkt), LMU Plus (unbeschränkt), LMU Profil (nur Lehrende), Student und Arbeitsmarkt (unbeschränkt).

- das persönliche Kennenlernen der Teammitglieder, in unserem Falle bei einer gemeinsamen mehrtägigen Exkursion;
- regelmäßige persönliche Treffen im Wochentakt; Arbeitsfortschritte, Ziele, sachliche und soziale Probleme und Lösungsansätze ließen sich so am besten kommunizieren.

Nicht bewährt hat sich, vor allem bei der Kommunikation von Aufgaben, die ‚gute alte‘ E-Mail; in einem Parallelprojekt hat sich auch ein Weblog nicht bewährt.

Neben der ‚Arbeit am Team‘ nimmt Projektmanagement eine herausgehobene Position ein. Für die Motivation der Studierenden und einen erfolgreichen Projektabschluss ist es wichtig, Projekte den universitären Rahmenbedingungen entsprechend zu dimensionieren. So sollten Projekte den Zeitraum eines Semesters nicht überschreiten. Allerdings hat sich in anderen geschichtswissenschaftlichen Lehrveranstaltungen, in denen besondere IT-Kenntnisse wie PHP/MySQL am konkreten Forschungsgegenstand vermittelt wurden, gezeigt, dass ein Semester kaum ausreicht. Im Zweifel sind die Projekte daher lieber „klein“ zu dimensionieren.

Weiter sind Lösungen für das Problem der Ausweisung individueller Prüfungsleistungen innerhalb von Gemeinschaftsprojekten zu finden. Aus studentischer Perspektive sind prüfungsadäquate Leistungen oft schon durch die notwendigerweise besonders intensive Mitarbeit gegeben, etwa durch die Eingabe von Daten. Insbesondere dann, wenn als Projektziel die Option Publikation gewählt wird, ist damit umzugehen, dass Datenbereinigung (insbesondere die Kontrolle der Transkription) mindestens ebenso viel Zeit verbraucht wie die sorgfältige Datenaufnahme (Transkription). Der Faktor 2 scheint hier realistisch; rechnet man die ‚Publikationszielgerade‘ hinzu, sollte man mit dem Faktor 3 rechnen. Wird dies am Anfang der Projektarbeit klar kommuniziert, bleibt die Motivation bei den Studierenden erhalten – es bewährt sich dann die „Impftheorie“ (Dieter Frey).

3. Fazit und Ausblick

Der Online-Editor *Squirrel* ist ein Editionstool für Lehre und Forschung. Es basiert auf der interdisziplinären Zusammenarbeit von Geschichtswissenschaft, Informatik und Computerlinguistik. Er ermöglicht den Studierenden individuelles online-basiertes Lernen und verbindet dieses optimal mit Publikationsprojekten. Das Tool wurde speziell für den Einsatz für Editionsprojekte in der Lehre entwickelt. Es erfasst alle Projektphasen vom Basisniveau (als wissenschaftliche Übung) bis hin zur Publikation mit Druck- und Onlineschnittstelle (LaTeX/PDF, XHTML).

Durch die Implementation von Funktionen aus UNITEX/GramLab werden eine neuartige Registererstellung und bessere Durchsuchbarkeit der Texte ermöglicht (Disambiguierung, Lemmatisierung, Konkordanz) sowie Optionen für eine semantische Suche angeboten. Da es möglich ist, für neue Projekte auch das visuelle Vergleichsmaterial früherer Projekte verfügbar zu machen, vergrößert sich mit jedem neuen Projekt die Menge des Vergleichs- und Übungsmaterials.

<http://www.squirrel.uni-muenchen.de/>