

Semantische Minimal-Retrodigitalisierung von Brief-Editionen

Rettinghaus, Klaus

klaus.rettinghaus@gmail.com

Sächsische Akademie der Wissenschaften zu Leipzig,
Deutschland

Situation

Brief-Editionen sind in vielerlei Hinsicht ein unerlässliches Hilfsmittel für die (historische) Forschung. Gerade weil Briefe ein höchst subjektives Ausdrucksmedium sind, enthalten sie oftmals wertvolle Informationen, die für das Verständnis größerer Zusammenhänge essentiell sein können. Doch um ihr volles Potential entfalten zu können, müssten sie leicht zugänglich sein. Dabei liegen die allermeisten Editionen ausschließlich in gedruckter Form vor, nur neuere Vorhaben publizieren wenigstens teilweise digital.

Der Forscherin oder dem Forscher entgehen im Zweifelsfall wichtige Schriftzeugnisse, weil sie sie schlichtweg nicht finden. Oder um es mit Goethe zu sagen: „Man erblickt nur, was man schon weiß und versteht.“ Es stellt sich also die Frage, wie man diese zwischen Buchdeckel gepressten Schätze heben und der Forschung zugänglich machen kann.

OCR-gestützte Retrodigitalisierung von Volltexten kritischer Editionen ist hier qualitativ zumeist noch nicht ausreichend. Erschwerend kommt hinzu, dass diese nicht-semantic ist – und bei urheberrechtlich geschützten Werken in aller Regel völlig unmöglich.

Lösungsansatz

Als möglicher Lösungsansatz wäre eine sogenannte semantische Minimal-Retrodigitalisierung vorzuschlagen, um gedruckte Editionen zu erschließen und online auffindbar zu machen. Hierbei werden ausschließlich die Metadaten der Briefe digitalisiert, das heißt die Namen des Schreibers und des Empfängers sowie das Datum des Briefes. Weitere Informationen, beispielsweise zu den Orten, sind wünschenswert, jedoch nicht notwendig. Versieht man diese Metadaten mit Normdaten und stellt sie im TEI-Austauschformat CMIF (*Correspondence Metadata Interchange-Format*) bereit, könnten sie auch im Kontext von vielen tausend weiteren Briefen im Webservice *correspSearch.net* sichtbar werden, der von der Berlin-Brandenburgischen Akademie der Wissenschaften bereitgestellt wird.

Der große Vorteil hierbei ist, dass auch jüngste Publikationen erschlossen werden können, da auf den notwendigen Metadaten kein Urheberrechtsschutz liegt. Nicht selten enthalten solche Editionen sogar ein tabellarische Korrespondenzregister, das relativ einfach gescannt und aufbereitet werden kann. Generell scheint es allerdings nicht ratsam die Brief-Metadaten direkt in XML zu kodieren, da ein solches Vorgehen relativ fehleranfällig und zeitaufwändig ist. Aufgrund des hohen Grades an Strukturiertheit ist es sinnvoller sie gleichsam in Tabellen zu erfassen; dies erhöht die Übersichtlichkeit und reduziert die Fehlerquote. Wünschenswert ist natürlich eine Anreicherung mit Normdaten. Diese kann manuell oder auch halbautomatisch mit entsprechenden Tools erfolgen. Die kompilierten Daten können dann mit dem Programm *CSV2CMI* automatisch in das entsprechende TEI-XML umgewandelt werden. Die resultierende Datei muss dann nur noch im World Wide Web bereitgestellt werden.

Ein weiterer, charmanter Vorteil hierbei ist, dass in der Tabelle beliebig viele zusätzliche Informationen untergebracht werden können, die im CMIF nicht oder noch nicht kodiert werden können. So besteht folglich sogar die Möglichkeit zwei Dateien bereitzustellen, eine für den automatisierten und standardisierten Austausch (*machine-readable*) und eine für Endbenutzer leicht lesbare Übersicht (*human-readable*).

Ausblick

Gleichwohl das hier vorgestellte Verfahren vergleichsweise unkompliziert ist, macht sich die Arbeit dennoch natürlich nicht von allein. Für wen erscheint es also sinnvoll diese Daten aufzubereiten und bereitzustellen?

Zunächst kommen natürlich Bibliotheken in den Sinn, die – als Informations-Anbieter des 21. Jahrhunderts – ein höchst eigenes Interesse daran haben (sollten), Wissen zu und aus Ihren Beständen anzubieten. Des Weiteren kommen Forscher in den Sinn, die sowieso intensiv mit einem Korrespondenzkorpus arbeitet. Das Eigeninteresse an verbesserter Auffindbarkeit und Zugänglichkeit dort sollte ausreichen, die notwendigen Daten zu digitalisieren und vorrätig zu halten. Hinzu kommt in letzterem Falle natürlich die Möglichkeit, die so aufbereiteten Daten als Forschungsdaten zu publizieren.

Zusammenfassung

Das Poster soll die hier skizzierte Idee der Semantischen Minimal-Retrodigitalisierung von Brief-Editionen vorstellen und den vorgeschlagenen Workflow sowie die bereit stehenden Software-Tools präsentieren.

Bibliographie

Woesler, Winfried (1988): *Vorschläge für eine Normierung von Briefeditionen*, in: *Editio* 2, S. 8–18. doi: 10.1515/9783110241938.8

Ball, Rafael (2014): *Bibliotheken im 21. Jahrhundert. Vom Leser zum Kunden*, in: **Ceynowa, Klaus / Hermann, Martin:** *Bibliotheken: Innovation aus Tradition*, Berlin: De Gruyter Saur 226–231, doi: 10.1515/9783110310511.226

Stadler, Peter (2014): *Interoperabilität von digitalen Briefeditionen*, in: **Wolzogen, Hanna Delf von / Falk, Rainer:** *Fontanes Briefe ediert*, Würzburg: Königshausen & Neumann 278–287.

Dumont, Stefan (2016): *correspSearch – Connecting Scholarly Editions of Letters*, in: *Journal of the Text Encoding Initiative* 10, <https://journals.openedition.org/jtei/1511> [letzter Zugriff 10. Oktober 2018].

Rettinghaus, Klaus (2018): *saw-leipzig/csv2cmi* (Version v1.6.2), Zenodo, doi: 10.5281/zenodo.1461642 [letzter Zugriff 13. Oktober 2018].