

# Semi-automatische Erschließung von Rechnungsbüchern am Beispiel des Stadtarchivs Leuven

**Bigalke, Jan**

jbigalke@smail.uni-koeln.de  
Universität zu Köln, Deutschland

**Drach, Sviatoslav**

s.drach@uni-koeln.de  
Universität zu Köln, Deutschland

**Neuefeind, Claes**

c.neuefeind@uni-koeln.de  
Universität zu Köln, Deutschland

Historische Rechnungsbücher stellen eine reichhaltige Quelle für geschichtswissenschaftliche Fragestellungen dar. Sie geben einen detaillierten Einblick in die Alltagsgeschichte, indem sie offenlegen, was in welchem Zeitraum gehandelt wurde und von wem.<sup>1</sup> So können bspw. Überschüsse und Mängel im Warenverkehr rekonstruiert, Handelsbeziehungen nachgezeichnet oder auch Rückschlüsse auf die Lebensumstände in bestimmten Zeiten gezogen werden. Z.B. lassen sich aus klösterlichen Rechnungsbüchern Informationen über die Besitzungen eines Klosters gewinnen, etwa wie viele Hörige das Kloster zu einer bestimmten Zeit hatte und welche Abgaben diese zu leisten hatten (siehe z.B. Bruch 2013, Lübbers 2009). Vergleicht man dann die verzeichneten Abgabemengen über Zeiträume hinweg, können so bspw. reiche Ernten oder Missernten identifiziert werden (Gleba 2016: 53ff).

In unserem Beitrag präsentieren wir einen Prototyp zur semi-automatischen Erschließung einer großen Sammlung von Rechnungsbüchern des Stadtarchivs Leuven, den wir derzeit im Kontext des Projekts “Itinera Nova” entwickeln.<sup>2</sup> Im Rahmen des Projekts, das sich seit 2009 der Erschließung der im Stadtarchiv der belgischen Stadt Leuven gesammelten 1127 Schöffenregister aus den Jahren 1362-1795 mit einem Umfang von ca. 1.000.000 Seiten widmet, wurden seit 2019 auch die dort ebenfalls vorliegenden 457 Rechnungsbücher digitalisiert. Die Erschließung der Schöffenregister wird derzeit von einer Community mit rund 50 Freiwilligen betrieben und von Archiv-MitarbeiterInnen gemanagt. Das Cologne Center for eHumanities (CCeH) stellt dabei u.a. eine spezifisch für das Projekt entwickelte Web-Plattform für die kollaborative manuelle Erschließung der Schöffenregister bereit, die Rechnungsbücher hingegen sind bislang noch nicht erschlossen.

Um mit den in Rechnungsbüchern enthaltenen Informationen auch quantitativ arbeiten zu können, müssen sie in strukturierter Form erfasst werden.<sup>3</sup> Dabei müssen die enthaltenen Kennzahlen nicht nur transkribiert, sondern auch nach ihrem jeweiligen Typ unterschieden werden, etwa nach verschiedenen Ausgabe- und Einnahmearten, etc. Nicht zuletzt aufgrund des sehr großen Umfangs von mehreren hunderttausend Seiten ist eine vollständige

manuelle Transkription kaum zu leisten. Jedoch können Methoden der Digital Humanities hier eine starke Unterstützung bieten. Dabei lässt sich der Umstand nutzen, dass Rechnungsbücher im Vergleich zu anderen historischen Quellen über einen relativ hohen Grad an Strukturierung verfügen (Pollin 2020: 5ff), insbesondere bei Kostenaufstellungen, die in Form von Tabellen festgehalten sind. Hierbei denkt man in der Regel zuerst an vorgedruckte bzw. vorgezeichnete Tabellen mit sichtbaren Linien. Neben diesen gibt es aber auch Rechnungsbücher, die rein konzeptionelle Tabellen ohne diese Linien beinhalten - so auch die Rechnungsbücher des Stadtarchivs Leuven. Ein Workflow zur Informationsextraktion aus Rechnungsbüchern muss demnach neben der Digitalisierung und Transkription auch eine Erkennung der Tabellen und eine anschließende Entity Recognition beinhalten. Für die Transkription existieren schon diverse etablierte Services wie beispielsweise Transkribus.<sup>4</sup> Diese bieten teilweise auch bereits Lösungen zur Layouterkennung, die auf Tabellen mit Linien ausgerichtet sind. Rein konzeptionelle Tabellen können dagegen nicht effektiv mit diesen Tools erkannt werden.

Für diese Problematik wurde von uns ein Prototyp entwickelt, mit dem auch konzeptionelle Tabellen automatisch erfasst werden können. Genutzt werden hierfür die Koordinaten der einzelnen Zeilen, wie sie auch für die automatische Texterkennung erfasst werden müssen. Die Koordinaten müssen hierbei in XML im PAGE-Schema<sup>5</sup> vorliegen. Erzeugt werden können diese Daten zum Beispiel durch eine Layoutanalyse mit dem Tool P2PaLA.<sup>6</sup> Bei den Tabellenzeilen wird davon ausgegangen, dass sie mindestens zweispaltig sind und dass die Spalten einen messbaren Abstand zueinander haben. Daher wird zur Erkennung der einzelnen Zeilen ein speziell trainiertes P2PaLA Modell genutzt, das die zwei Spalten einer Tabellenzeile als zwei einzelne Zeilen erkennt. Das Tool vergleicht hierfür sämtliche benachbarten Textzeilen miteinander. Zuerst werden die Positionen der Textzeilen auf der y-Achse miteinander verglichen. Unterschreitet die Distanz der Textzeilen einen bestimmten Wert, wird davon ausgegangen, dass die Textzeilen auf der gleichen Zeile stehen. Als nächstes werden nun die Positionen der Textzeilen auf der x-Achse miteinander verglichen. Wird hier eine festgelegte Differenz überschritten, ist dies ein Zeichen dafür, dass es sich bei den Textzeilen um zwei Spalten einer Tabelle handelt. Wiederholt sich dieses Phänomen, ist dies ein starker Hinweis darauf, dass es sich bei den Textzeilen um Zeilen einer Tabelle handelt.

Dieses Vorgehen stellt eine robuste Methode dar, um rein konzeptionelle Tabellen eines Rechnungsbuches zu erkennen. Durch das Anpassen der Regeln ist dieses Vorgehen nicht nur auf Rechnungsbücher aus einer Quelle beschränkt, sondern kann leicht an die Gegebenheiten anderer Rechnungsbücher angepasst werden.

## Fußnoten

1. Unter anderem beschreibt Paul Kirn in seiner Einführung in die Geschichtswissenschaften Quellen zur Wirtschaftsgeschichte als typisch mittelalterliche Quellenart (Kirn 2019: 33).
2. Das Projekt Itinera Nova ist über den Link <https://www.itineranova.be> zu erreichen. Dort können auch die bereits digitalisierten Objekte eingesehen werden.
3. Vgl. dazu v.a. auch die Überlegungen von Georg Vogeler (2015) zu digitalen Rechnungsbüchern.
4. Siehe <https://readcoop.eu/transkribus>
5. PAGE steht für “Page Analysis and Ground-Truth Elements” (Pletschacher/Antonacopoulos 2010).

6. P2PaLA steht für “Page to PAGE Layout Analysis”, siehe <https://github.com/lquiroso/P2PaLA>. Eine Dokumentation findet sich bei Quiròs (2018).

## Bibliographie

**Bruch, Julia** (2013): Die Zisterze Kaisheim und ihre Tochterklöster. Studien zur Organisation und zum Wirtschaften spätmittelalterlicher Frauenklöster mit einer Edition des Kaisheimer Rechnungsbuches. Berlin.

**Gleba, Gudrun** (2016): Rechnen. Wirtschaften. Aufschreiben. Vernetzte Schriftlichkeit - Wirtschaft und Rechnungsbücher als Quellen klösterlicher Alltagsgeschichte. In: Pätzold, Stefan und Stumpf, Marcus (Hg.): Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster. S. 51-64.

**Kirn, Paul / Leuschner, Joachim** (2019): Einführung in die Geschichtswissenschaften. Berlin / Boston: De Gruyter.

**Lübbbers, Bernhard** (2009): Die Ältesten Rechnungen des Klosters Alderbach (1291 - 1373/1409). Analyse und Edition. In: Quellen und Erörterungen zur bayerischen Geschichte: NF 47.3. München: Beck.

**Pletschacher S. / Antonacopoulos A.** (2010): The PAGE (Page Analysis and Ground-truth elements) Format Framework”, in: ICPR2010, Istanbul. S. 257-260.

**Pollin, Christopher** (2020): Digitale, formale Methoden und Modelle in den Geschichtswissenschaften. Am Beispiel digital editierter historischer Rechnungsbücher. [https://chpollin.github.io/paper/Pollin\\_MA\\_2020\\_Geschichte.pdf](https://chpollin.github.io/paper/Pollin_MA_2020_Geschichte.pdf) [letzter Zugriff 14. Juli 2021].

**Quiròs, Lorenzo** (2018): Multi-Task Handwritten Layout Analysis <https://arxiv.org/pdf/1806.08852.pdf> [letzter Zugriff 14. Juli 2021].

**Vogeler, Georg** (2015): Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? In: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum / Thomas Stäcker (Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). DOI: 10.17175/sb001\_007.