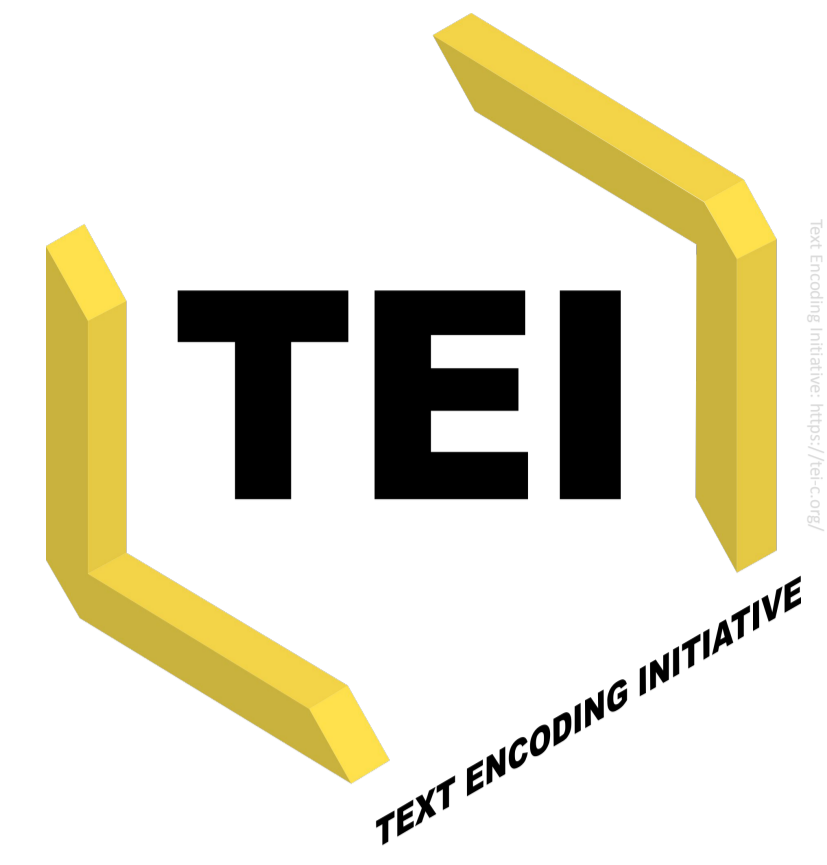
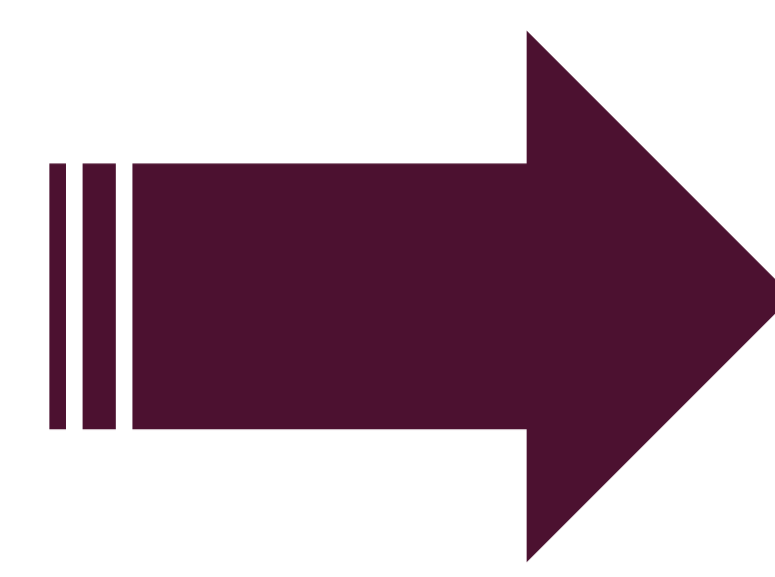
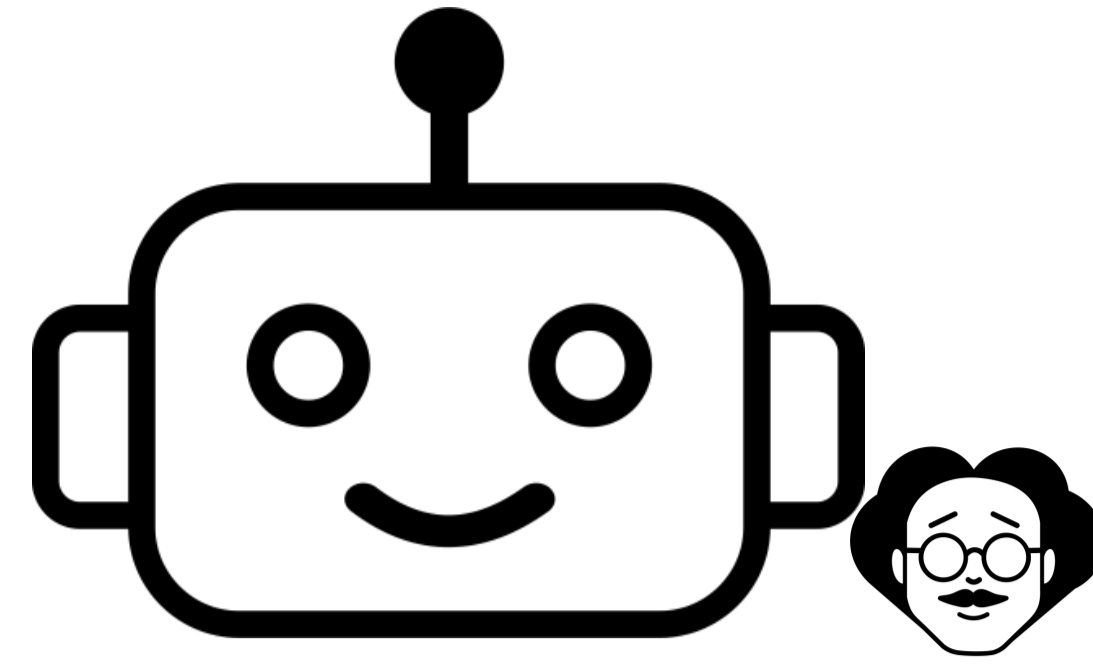
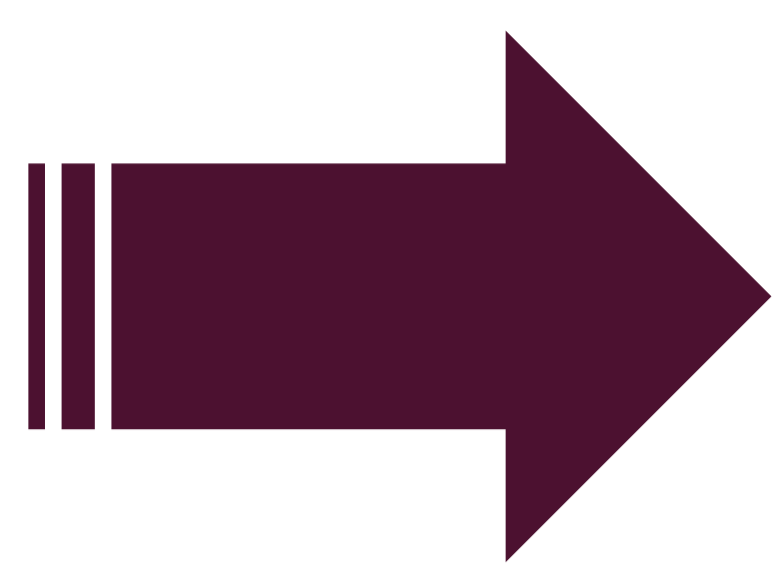
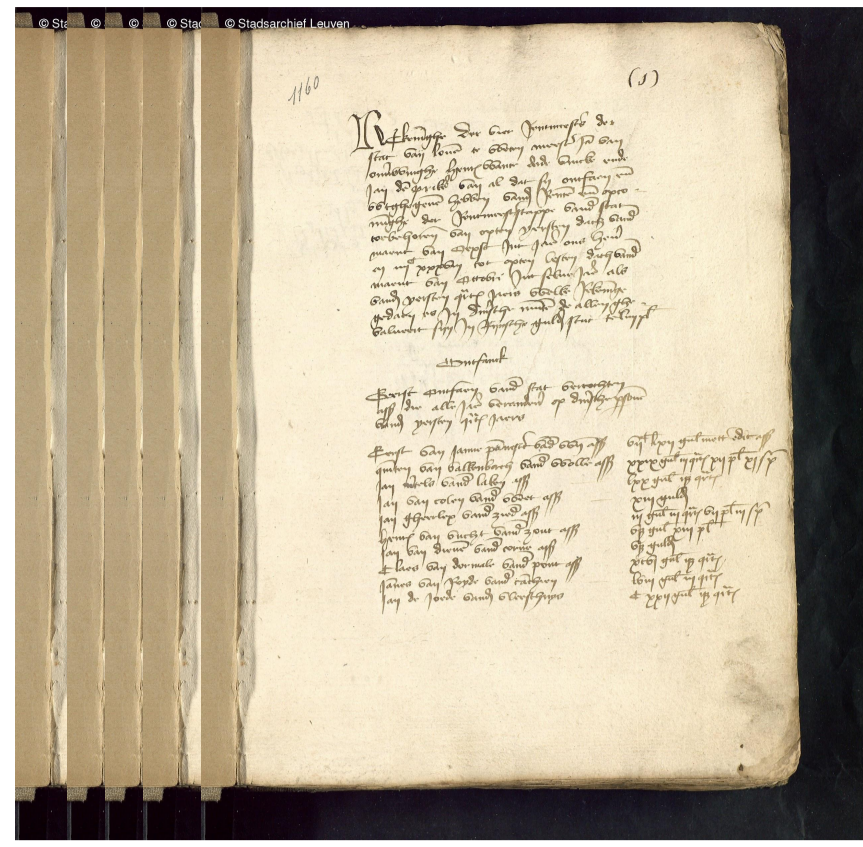


Semi-automatische Erschließung von Rechnungsbüchern am Beispiel des Stadtarchivs Leuven

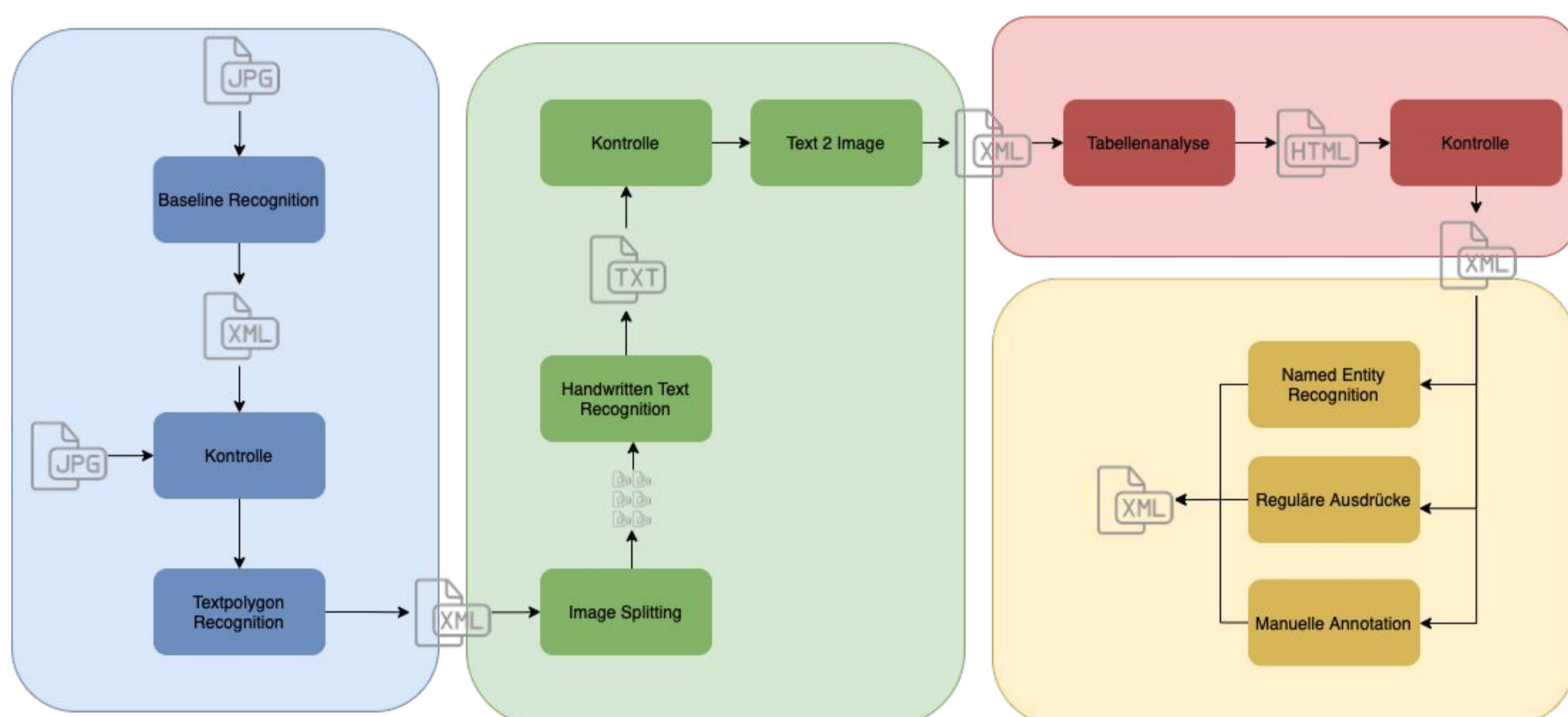


Worum geht es?

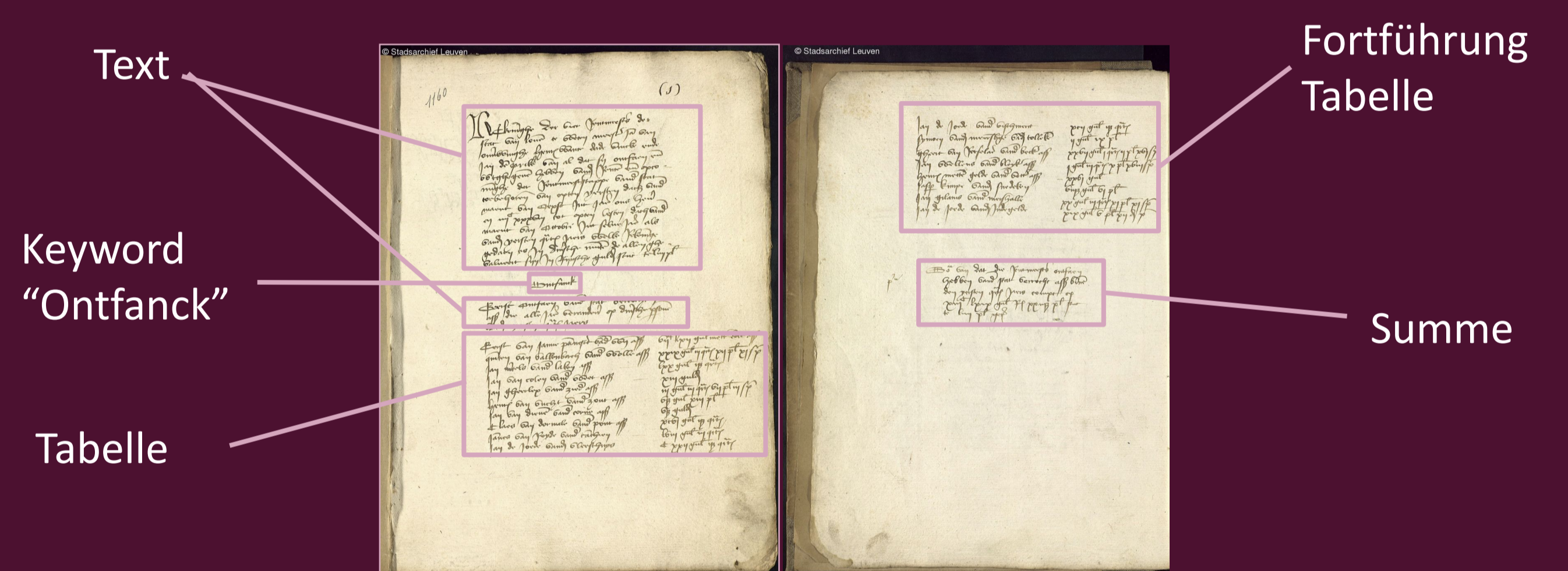
In unserem Poster präsentieren wir einen Prototyp zur semi-automatischen Erschließung einer großen Sammlung von Rechnungsbüchern des Stadtarchivs Leuven, den wir derzeit im Kontext des Projekts "Itinera Nova" entwickeln. Im Rahmen des Projekts, das sich seit 2009 der Erschließung der im Stadtarchiv der 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.

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. Der Workflow beschreibt das Vorgehen ausgehend vom Digitalisat bis hin zur ausgezeichneten XML, einschließlich der eingesetzten Tools, der Input- und Output-Daten, sowie den nötigen Transformationsschritten.

Workflow

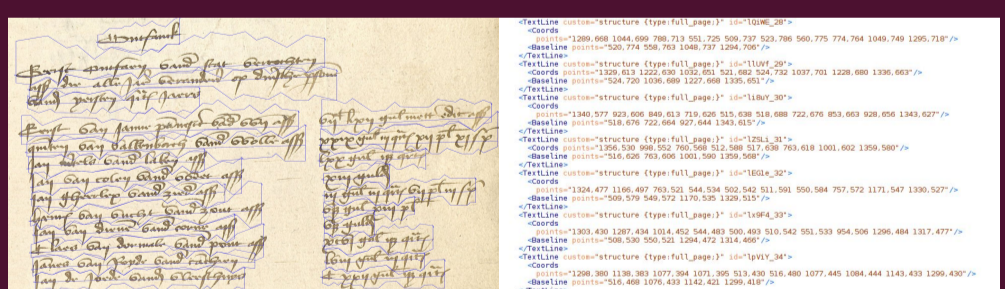


Bestandteile der Rechnungsbücher

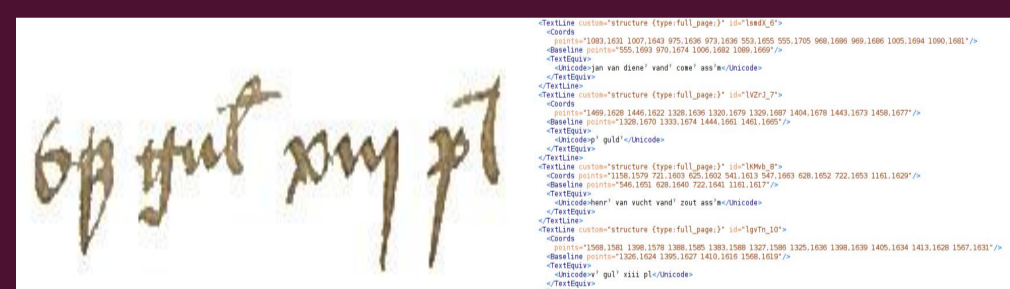


Prozessschritte

1. Die Konturen des Textes innerhalb des Digitalisats werden mit P2PaLA erkannt und in einem PAGE-XML gespeichert.



2. Mithilfe eines HTR-Verfahrens (PyLaia) wird der Text des Digitalisats erkannt.



3. Die Abstände der einzelnen Zeilen zueinander werden verglichen. Unter Beachtung bestimmter Signalwörter können so die Tabellen erkannt werden.



4. Im Anschluss an die Tabellenanalyse können Annotationen hinzugefügt werden.



Fazit:

Das regelbasierte Vorgehen stellt eine robuste Methode dar, um rein konzeptionelle Tabellen in Rechnungsbüchern des Stadtarchivs Leuven zu erkennen. Durch das Anpassen der Skripte ist dieses Vorgehen nicht nur auf Rechnungsbücher aus einer Quelle beschränkt, sondern kann leicht auf die Gegebenheiten anderer Rechnungsbücher übertragen werden.

