

Forschungsprojekt zur automatisierten Handschriftenerkennung soll Zugang zu historischen Archivdokumenten revolutionieren

Pressemitteilung

Mit 8,2 Millionen EUR wird das Projekt READ (Recognition und Enrichment of Archival Documents) von der Europäischen Kommission im Rahmen des H2020 Programms gefördert. Forschung und Technologie „Made in Germany“ spielt dabei eine zentrale Rolle. So sollen die automatisierte Handschriftenerkennung, die Layout-Erkennung komplexer Dokumente, aber auch verbesserte Suchalgorithmen und die automatische Schreibererkennung die Erschließung von Archivdokumenten auf eine völlig neue Ebene heben. Zusätzlich zu einzelnen Technologiepaketen werden diese im Rahmen einer Serviceplattform frei zugänglich gemacht. Ein erster Prototyp zur Handschriftenerkennung steht bereits auf der Website der Plattform (<http://transkribus.eu/>) für jedermann frei zur Verfügung.

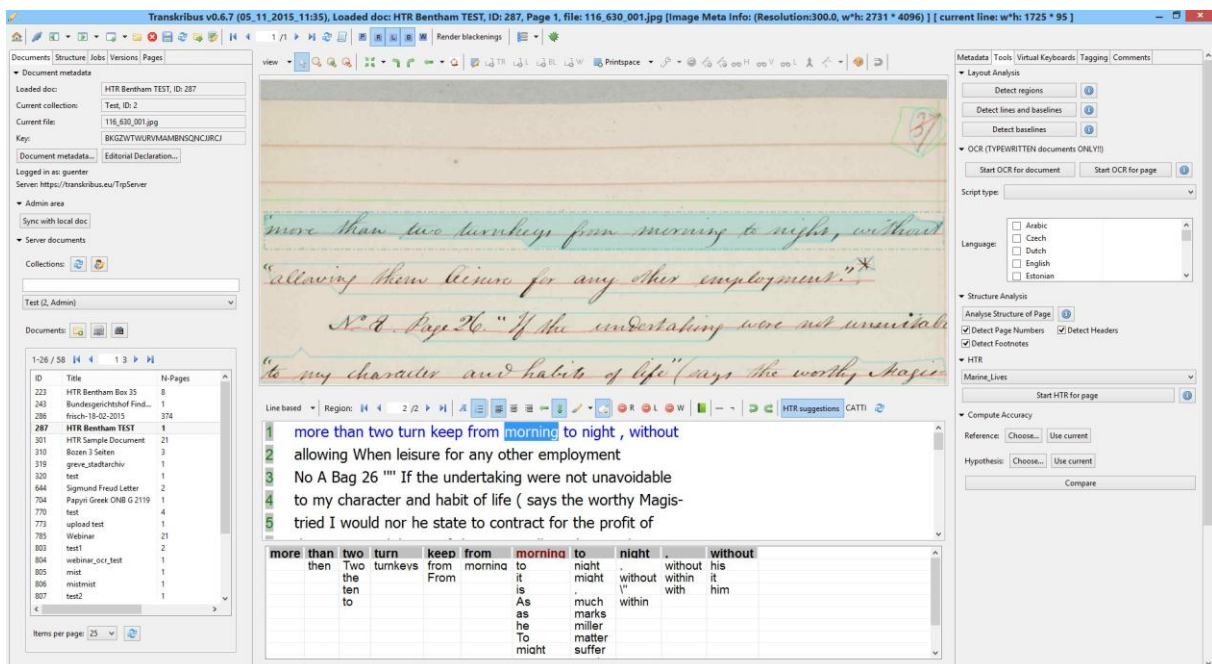
Viele hundert Millionen Akten, Urkunden, Rechnungsbände und Protokolle, Karten und Pläne aus mehr als 1200 Jahren europäischer Geschichte werden in den Archiven Europas verwahrt. Sie bilden das Gedächtnis Europas, seiner Staaten und Städte, seiner Gemeinden und seiner Bürgerinnen und Bürger, sie sichern die Rechte des einzelnen und machen das Verwaltungshandeln transparent.

Bei diesem Archivgut handelt es sich um einen der letzten „ungehobenen Schätze“ des kulturellen Erbes: Einerseits finden sich darin fast ausschließlich Unikate und andererseits wurde der allergrößte Teil dieser Dokumente bisher nicht veröffentlicht. Hinzu kommt, dass ein großer Teil der Dokumente handgeschrieben vorliegt, so daß lediglich Experten sie ohne Schwierigkeiten entziffern können. Nicht nur großen Teile der Bevölkerung sondern auch der Forschung bleiben daher viele der darin enthaltenen Informationen, Geschichten und Geheimnisse verborgen bzw. haben nicht jene Bedeutung in Forschung, Lehre und Öffentlichkeit, die ihnen eigentlich zukommt.

Immer mehr Archive gehen nun allerdings dazu über, Teile ihrer Bestände systematisch zu digitalisieren. Denn mit der Digitalisierung kann nicht nur der Zugriff auf die Archivdokumente unabhängig von Zeit und Raum gegeben werden, die Digitalisierung eröffnet auch die Chance, mittels modernster Methoden diese Dokumente völlig neu zu erschließen. Eine wichtige Rolle spielt dabei die automatische Handschriftenerkennung, die – nach vorhergehendem Training – bereits heute historische Dokumente mit überraschend hoher Genauigkeit von 70 bis 80% entziffern kann. Die Universität Rostock sowie die Technische Universität Valencia, beide Partner im READ-Projekt, gehören hier zu den führenden Forschungsgruppen. Aber auch die automatisierte Schreibererkennung, die nur aufgrund des „Schriftbildes“ einer Handschrift in der Lage ist, Dokumente einem bestimmten Schreiber zuzuweisen, bzw. diese nach diversen Kriterien wie Entstehungszeit oder Schriftart zu ordnen, bietet für den Benutzer völlig neue Möglichkeiten. „Gib mir alle Dokumente in der Handschrift Goethes innerhalb dieses Archivs“ könnte etwa eine

künftige Suchabfrage lauten. So arbeiten die Universität Erlangen sowie die Technische Universität Wien in diesem Bereich. Viele Archivadokumente weisen ein komplexes Layout auf, Tabellen und Formulare werden seit dem 16. Jahrhundert immer häufiger verwendet. Historiker und Sozialwissenschaftler, die historische Steuerabgaben oder die Passagierlisten von Auswandererschiffen im großen Umfang quantitativ untersuchen wollen, stoßen hier bisher rasch an die Grenzen. Eine verbesserte Layout-, Tabellen- und Formularerkennung, im Projekt durch mehrere Forschungsgruppen abgedeckt, soll hier Abhilfe schaffen. Im Projekt werden Archivadokumente aus diversen deutschen Institutionen verarbeitet, zu nennen sind hier nicht nur das Staatsarchiv Marburg, sondern auch das Archiv der Diözese Passau, der Gottfried Wilhelm Leibniz Bibliothek, der Forschungsbibliothek Gotha sowie des Musikinstrumenten-Museums der Universität Leipzig. Darüber hinaus sind auch das Nationalarchiv aus Finnland und andere internationale Archive im Projekt mit ihren Beständen vertreten. Koordiniert wird das READ-Projekt von der Universität Innsbruck.

Die Grundlagenforschung in den oben genannten Bereichen ist bereits weit fortgeschritten. Nun geht es darum, diese Forschungsergebnisse auch für Archive und Bibliotheken, Geisteswissenschaftler und Familienforscher sowie eine breite Öffentlichkeit nutzbar zu machen. Deshalb ist der Aufbau einer freien Serviceplattform von zentraler Bedeutung im READ-Projekt. Hier können die Technologien und Algorithmen erprobt werden, hier können aber auch Datensets erstellt werden, die von eminenter Bedeutung für die weitere Entwicklung der Technologie sind. Denn die allermeisten der angewandten Verfahren beruhen auf maschinellen Methoden, und diese benötigen Daten. Je mehr Dokumente in der Plattform bearbeitet werden, desto mehr Daten stehen der Forschung zur Verfügung und desto besser sollten künftige Modelle funktionieren. Viele der eingesetzten Verfahren sind sprachunabhängig. Die Forschungsplattform richtet sich daher nicht nur an deutschsprachige Benutzer, sondern ist von vornherein international ausgerichtet. Eine Teilnahme mit eigenen Dokumenten ist sowohl für Geisteswissenschaftler, Archive aber auch Privatpersonen jederzeit möglich und gewünscht.



Informatiker und Archivare im Dienst an der gemeinsamen Sache

Aber das Projekt braucht nicht nur Informatiker für die Weiterentwicklung der Software, sondern auch Spezialisten für die alten Dokumente und Schriften. So entstand eine Kooperation mit einem weiteren von der Europäischen Union geförderten Projekt, das sein Ziel schon im Namen trägt: „co:op – Community as Opportunity. The creative Archives' and Users Network“. Es handelt sich um einen Zusammenschluß von 17 Archiven und Universitäten aus 11 Ländern unter der Leitung von Dr. Francesco Roberg vom Hessischen Staatsarchiv Marburg. Die hier organisierten Archivare und Historiker suchten ihrerseits nach einem Weg, große Mengen von einzigartigen, aber nur schwer lesbaren Archivalien einer möglichst großen Öffentlichkeit nahezubringen. Das READ-Projekt stellt folglich die ideale Ergänzung dar. Auf diese Weise werden die Kompetenzen beider Projekte – Informatiker einerseits und Archivaren andererseits, die mit den alten Dokumenten umgehen können und über große Mengen digitalisierter Archivalien verfügen – gebündelt.

Beide Projekte tagen öffentlich vom 19.-21. Januar 2016 im Hessischen Staatsarchiv Marburg, Friedrichsplatz 15, 35037 Marburg (Technology meets Scholarship, or how Handwritten Text Recognition will Revolutionize Access to Archival Collections. With a special focus on biographical data in archives). Erwartet werden Teilnehmer aus mehr als 20 europäischen und außereuropäischen Ländern, darunter aus Ägypten, Israel, Marokko und den USA. Roberg zufolge würde eine erfolgreiche Weiterentwicklung des Programms bis hin zur vollautomatisierten Lesbarkeit handgeschriebener Texte einen Paradigmenwechsel bedeuten, dessen Konsequenzen noch gar nicht in vollem Umfang absehbar sind.

Mit der Software und der Unterstützung anderer Nutzer sollte es dann auch möglich sein, den in Kurrentschrift verfassten Brief des Großvaters rasch zu entziffern. So sollen in den nächsten Jahren Katasterbände, Kirchenbücher, Briefe, aber auch unterschiedlichste Personenlisten (Immigranten, Passagiere, etc.), Ratsprotokolle und viele andere historische Dokumente computerlesbar gemacht werden. Die Forscher werden dafür auch eine App für Smartphones anbieten, mit der die Handschriften direkt eingescannt werden können. Um die Menschen zum Mitmachen zu motivieren, sollen gemeinsam mit der Öffentlichkeit auch die Handschriften berühmter Persönlichkeiten gesammelt und automatisch erkennbar gemacht werden. Dazu zählt etwa der Nachlaß der weltberühmten Brüder Grimm im Hessischen Staatsarchiv Marburg. Alle diese digitalisierten Handschriften können dann am Computer durchsucht werden.

Weiterführende Hinweise:

1. Eine experimentelle Version der Software steht im Internet bereits zum Download zur Verfügung. Expertinnen und Experten, aber auch Laien können sich dort registrieren und das Programm ausprobieren: <http://ranskribus.eu>
2. Tagungsprogramm: <https://landesarchiv.hessen.de/aktuelles/veranstaltungskalender>
3. Homepage des co:op-Projekts: <http://coop-project.eu/> und des READ-Projekts: <http://read.transkribus.eu/>
4. Ansprechpartner co:op-Projekt und Tagung

Dr. Francesco Roberg
Hessisches Staatsarchiv Marburg

Friedrichsplatz 15
35037 Marburg
Tel.: +49 6421 9250-160
E-Mail: f.roberg@stama.hessen.de
Web: <https://landesarchiv.hessen.de/>

5. Ansprechpartner READ-Projekt

Mag. Dr. Günter Mühlberger (Koordinator READ)
Digitalisierung und elektronische Archivierung (DEA)
Universität Innsbruck
Tel.: [+43 512 507 8454](tel:+435125078454)
E-Mail: guenter.muehlberger@uibk.ac.at
Web: www.uibk.ac.at/germanistik/dea/

Mag. Dr. Christian Flatz
Büro für Öffentlichkeitsarbeit
Universität Innsbruck
Tel.: [+43 512 507 32022](tel:+4351250732022)
E-Mail: christian.flatz@uibk.ac.at