

Automatische Klassifikation unstrukturierter multilingualer Dokumente – Mögliche Lösungsszenarien und Szenarienwahl

Ruth Zürcher
laruz@bluewin.ch

Obwohl die Mitarbeitenden des Eidgenössischen Departements für auswärtige Angelegenheiten (EDA) seit 2009 verpflichtet sind, Dokumente in einem Geschäftsverwaltungssystem (GEVER-System) abzulegen, sind bis 2018 auf verschiedenen Fileshares rund 12 Mio. Dokumente entstanden. Die Bestände sind archivisch nicht bewertet und grösstenteils nicht gemäss dem offiziellen Registraturplan strukturiert.

Sie sollen nicht mehr weiter anwachsen, sondern gemäss dem offiziellen Registraturplan strukturiert, ins GEVER-System überführt und dem Schweizerischen Bundesarchiv (BAR) zur Archivierung angeboten werden.

Die vorliegende Masterarbeit geht der Frage nach, inwiefern es möglich wäre, die Zuordnung der Dokumente zu den passenden Rubriken des neuen Registraturplans durch informationslinguistische Tools zu unterstützen, und welche Lösungsszenarien es dafür gibt.

Zuerst wird kurz dargestellt, wie die heutige Situation entstand, und welche spezifischen Probleme sich im EDA stellen. Danach werden verschiedene Methoden der maschinellen Textverarbeitung und die Arbeitsschritte der automatischen Textklassifikation vorgestellt. Anschliessend werden die Rahmenbedingungen aufgezeigt, welche elektronische Werkzeuge innerhalb des EDA erfüllen müssten, um zum Einsatz zu kommen. Weiter stellt sich die Frage, ob die benötigten Hilfsmittel für eine maschinelle Zuordnung vorhanden sind, resp. wo sie erhältlich wären.

Um herauszufinden, ob geeignete Tools bereits kommerziell verfügbar sind und um ihr Potential im vorliegenden Fall zu ermitteln, wurde mit fünf Firmen ein Request for Information (Rfi) durchgeführt, der ergab, dass solche Produkte am Markt erhältlich sind.

Zum Schluss werden zwei mögliche maschinelle Vorgehensszenarien und die manuelle Zuordnung miteinander verglichen. Ihre jeweiligen Vor- und Nachteile werden beleuchtet, sie werden aufgrund fachlicher, technischer und projektbezogener kritischer Erfolgsfaktoren beurteilt und eine Umsetzungsempfehlung abgegeben.