Presentation

Aaron Richiger

Automated Data Extraction from Documents using Machine Learning

Abstract:
About 80 percent of all business-relevant data is unstructured. Given the variety in layouts and changing contexts of information in a document, data fields or texts are still manually copied for further processing. This is time-consuming, error-prone and costly. Bringing semi-structured and unstructured data into a machine-readable format is therefore an ongoing challenge and prerequisite also in the field of text analytics.
Over the last years, we developed a document extraction platform – MINT.extract – that allows us to access relevant data (text, images, etc.) from any type of document in a highly efficient and flexible way. In this talk, we present a customer case to process purchase orders using machine learning. Besides automated extraction of information like order number, article number, ordering date, and the number of items ordered, we also perform validation steps. The system runs as a service 24/7 and is accessible from all our client's global locations. As a further benefit, the learning system is highly scalable given the small set of training data required to get the accuracy scores above 95%. We will share some technical insights on how we achieved this. Finally, an intuitive user interface allows users to train their own learning system independently and the architecture enables a seamless integration with existing ERP systems. Naturally, this approach can be applied to insurance policies, invoices, official documents or entire book series as well as other document types.

Biography: Aaron Richiger is a passionate entrepreneur and gifted software engineer working full-time at turicode AG, which he co-founded. He employs his persistent scientific curiosity to develop novel software solutions and is responsible for Machine Learning within turicode. He completed both his bachelor's and master's degree in computer science at ETH Zurich.

Organization: turicode AG

Contact: aaron.richiger@turicode.com