

snapADDY GmbH mejora el reconocimiento de datos mediante el aprendizaje profundo

01 de febrero de 2019

Con la última actualización de sus productos, snapADDY GmbH, de Würzburg, Alemania, utiliza un nuevo analizador sintáctico que mejora aún más el reconocimiento de los datos de contacto gracias al aprendizaje profundo. Este método es el elemento clave de todos los productos (snapADDY Grabber, VisitReport, CardScanner y Assistant) para detectar los datos de contacto a partir de textos no estructurados y, por lo tanto, es el núcleo técnico de los productos snapADDY.

Funcionamiento del analizador sintáctico snapADDY

El analizador de datos snapADDY garantiza el reconocimiento de los datos contacto de diferentes fuentes de texto. Por ejemplo, puede tratarse de un texto simple impreso, o de un correo electrónico cuya firma contenga datos relevantes para snapADDY. Además, otro caso de uso del analizador snapADDY es la foto de una tarjeta de presentación, que primero se "lee" mediante el reconocimiento de texto. Para las máquinas, el reconocimiento de los datos de contacto no es una tarea trivial (a diferencia de los seres humanos) pues estas requieren de una mezcla de reglas complejas y una evaluación inteligente del contenido identificado. Para ello, se examinan individualmente los componentes de un contacto. Algunos datos de contacto, como las direcciones de correo electrónico o las URL, tienen un formato uniforme y, por lo tanto, el analizador sintáctico puede identificarlos fácilmente mediante reglas.

Mejora del analizador sintáctico snapADDY con técnicas de aprendizaje profundo

Además del método de reglas, snapADDY GmbH adopta un nuevo enfoque con el analizador sintáctico lanzado a principios de año: en el caso de los datos de contacto no estandarizados, el reconocimiento es cada vez más complejo y ya no se es posible asignar correctamente reglas. Esto afecta, por ejemplo, a los nombres y a las profesiones: El término "Zapatero" puede denotar una profesión, así como el apellido de una persona. Los seres humanos se limitan a deducir el significado correcto a partir del contexto. El uso del aprendizaje profundo (Deep Learning) permite que la máquina proceda de manera similar, asignando correctamente la información de contacto individual en función de los términos que le rodean. La información obtenida de este modo se combina en el analizador sintáctico snapADDY con la información obtenida de las reglas para lograr el mejor reconocimiento posible de los contactos.



Mayor calidad en el reconocimiento de contactos y datos

Con la actualización del analizador sintáctico lanzado a principios de enero, snapADDY GmbH ha conseguido mejorar notablemente la calidad del reconocimiento en sus productos. Para ser más específicos, en el caso de los campos de contacto no estandarizados, el aprendizaje profundo es de gran ayuda para obtener la información de contexto. Por ejemplo, el reconocimiento de la profesión o cargo ha mejorado alrededor de un 19% en comparación con la versión anterior del analizador sintáctico. Asimismo, se reconocen mejor los nombres y apellidos, con una mejora del 7% y los números de teléfono, con una mejora del 16% – como se muestra en el gráfico.