

Nächste Evolutionsstufe von Dokumenten mit „Semantic-Web“-Technologien

Auf der diesjährigen DOXNET Jahreskonferenz vom 20.-22. Juni 2022 hat Dr. Arthur Brack in einem Grundsatzvortrag einen Blick in die Zukunft gewagt und einen Ausblick auf die nächste Evolutionsstufe von Geschäftsdokumenten gegeben. Demnach werden maschinenlesbare Geschäftsdokumente den nächsten Innovationsschub in der digitalen B2B- und B2C-Kommunikation auslösen.



Vortrag über die Zukunft von Dokumenten auf der Doxnet

Aktuelle Herausforderungen der B2B- und B2C-Kommunikation

In den letzten Jahrzehnten haben moderne Informationstechnologien zur Digitalisierung diverser Industrien und Branchen verholfen. So wurden bspw. gebundene Produktkataloge durch Online-Shops und papierhafte Straßenkarten durch Navigationssysteme weitgehend vollständig ersetzt. Allerdings hat sich die Repräsentation von Geschäftsdokumenten in den letzten Jahrzehnten kaum verändert: So wird heute immer noch ein Großteil der Dokumente auf Papier gedruckt und physisch versendet. Es werden zwar mittlerweile immer mehr Dokumente auch elektronisch im PDF- oder HTML-Format zur Verfügung gestellt, allerdings sind die wesentlichen Informationen in den Dokumenten nach wie vor nicht maschinenlesbar (bis auf wenige Ausnahmen wie z. B. ZUGFeRD-Rechnungen).

Deshalb ist es nicht möglich Dokumente miteinander zu verknüpfen, bestimmte Informationen aus den Dokumenten für Anwendungen zu exportieren oder mit den

Dokumenten zu interagieren.

Der Digitalisierungsgrad von Geschäftsdokumenten entspricht somit heutzutage weitgehend „Produktkatalogen bzw. Straßenkarten in PDF-Dateien“.

Dementsprechend ist die B2C- und B2B-Kommunikation aktuell durch Medienbrüche und Verzögerungen oft sehr ineffizient. Eine Anreicherung von Geschäftsdokumenten mit maschinenlesbaren Informationen würde für Empfängern von Dokumenten eine Vielzahl von nützlichen Anwendungen ermöglichen und die B2B- und B2C-Kommunikation erheblich vereinfachen.

So könnten beispielsweise Rechnungsdaten automatisiert ins Buchhaltungssystem importiert werden und Privatpersonen könnten mit Dokumenten einfacher interagieren (auf Knopfdruck Rechnung bezahlen, per digitalem Formular antworten, Zählerstand eingeben, usw.) und zusätzliche Informationen erhalten (z. B. verknüpfte Dokumente wie AGBs, Informationen zu versicherten Objekten).

Nächste Evolutionsstufe: Maschinenlesbare Geschäftsdokumente

Mit Hilfe von „Semantic Web“-Technologien können Dokumente miteinander verknüpft und mit maschinenlesbaren Informationen angereicht werden. Diese Technologien existieren bereits seit über 20 Jahren und werden vom World Wide Web Consortium (W3C) standardisiert. Angewandt werden diese Technologien bspw. im Internet zur Optimierung von Suchmaschinen, zur maschinenlesbaren Strukturierung von Wissen in Wikipedia (s. wikidata.org) oder maschinenlesbare E-Mails (s. www.schema.org).

Auch in betrieblichen Informationssystemen werden Semantic-Web-Technologien zunehmend immer wichtiger (siehe Knowledge Graphs).

Um maschinenlesbare Geschäftsdokumente zu ermöglichen, werden zum einen passende Ontologien benötigt, mit denen die Inhalte von Dokumenten modelliert werden (z. B. Informationen einer KFZ-Versicherung). Zum anderen müssen Geschäftsdokumente um maschinenlesbare Informationen angereicht werden.

Die Entwicklung von standardisierten Ontologien ist eine große Herausforderung und muss in Zusammenarbeit mit Versendern und Konsumenten von Dokumenten für die entsprechenden Dokumentenarten erfolgen. Hierbei können bereits vorhandene Ontologien genutzt und erweitert werden. Die Anreicherung von Geschäftsdokumenten mit maschinenlesbaren Informationen kann für Dokumente im HTML-Format mit Hilfe JSON-LD und im PDF-Format in XMP-Headersätzen mit RDF erfolgen.

Diese Anreicherung kann in einem Output-Management-System einfach erfolgen, indem Rohdaten (z. B. XML-Daten für eine KFZ-Versicherungen), die für die Dokumentenformatierung benötigt werden, in die erforderlichen Formate (JSON-LD oder RDF) zu passenden Ontologien konvertiert und in die Dokumente eingebettet werden.

Fazit

Aber lohnt sich der eigene Aufwand für die Erarbeitung von Ontologien und die Anreicherung von Dokumenten? Die Antwort lautet eindeutig „ja“! Allein für die Output-Management-Abteilung ergibt sich zwar kein direkter Nutzen, aber durch diese Technologien werden deutliche Mehrwerte für Unternehmen und Kunden geschaffen. Neben einem Beitrag zur Nachhaltigkeit durch die Reduktion von physischen Dokumenten wird die B2B- und B2C-Kommunikation deutlich effizienter und Unternehmen werden als modern und innovativ wahrgenommen!

The screenshot shows a business document titled "Rechnung" (Invoice). It includes a thank you message for an order from May 5, 2018, and a table with columns: "Bezeichnung", "Stück", "Mit Vorsatz", "Ab Lager", and "Erfasst". Below the table are two buttons: "Mit PayPal bezahlen" and "Rückfrage per eMail". A separate section titled "Jährliche Zählerstandablesung" (Annual meter reading) addresses "Herr Kohlschreiber" and requests support for a meter reading by June 31, 2022. It includes a QR code and a button "Zählerstand erfassen" (Record meter reading).

So können z. B. Aktionen und Verknüpfungen für Geschäftsdokumente aussehen