

Dokumentenbasierte Steuerung von Geschäftsprozessen

Dominik Reichelt

Professur für Produktionswirtschaft und Informationstechnik

dominik.reichelt@iisys.de

Abstract: Geschäftsprozesse im Verwaltungs- und Dienstleistungsbereich werden häufig durch den Eingang von Dokumenten angestoßen. Hierfür ist es unerlässlich, dass sie den richtigen Mitarbeiter im Unternehmen oder der Organisation erreichen. Oftmals sind jedoch dem externen Sender die internen Organisationsstrukturen nicht klar, so dass eine zentrale Stelle angeschrieben wird. Diese muss dann das Dokument, basierend auf seinem *Inhalt*, an die zuständigen Kollegen weiterleiten. Dies kann beträchtlichen personellen Aufwand mit sich bringen. In der Forschungsarbeit wird ein System entwickelt, das diese Aufgabe maschinell erfüllen soll. Hierzu werden verschiedenartige Klassifikationsverfahren erprobt und hinsichtlich ihrer Verlässlichkeit beurteilt. Weiterhin werden Verbesserungen gegenüber gängigen maschinellen Verfahren angestrebt.

Einleitung

Gerade im Dienstleistungsbereich sind Dokumente der Dreh- und Angelpunkt vieler Geschäftsprozesse. Als offensichtliche Beispiele lassen sich öffentliche Verwaltung, Bildung oder die Versicherungsbranche anführen.

Für gewöhnlich bestimmt der Inhalt der Dokumente die konkreten Prozesse, die sie durchlaufen. Ein Brief an eine Versicherung wird grundlegend anders behandelt, wenn es sich um eine Schadensfallschilderung handelt, als wenn er sich auf eine Krankheit bezieht. Das gleiche Prinzip lässt sich auch auf die Behandlung von Formularen anwenden. Im Idealfall ist es möglich, einen großen Teil solcher Abläufe zu automatisieren.

Zielstellung, Forschungsfragen

Das Hauptziel der Forschungsarbeit besteht darin, verschiedene Ansätze zu demonstrieren, wie sich Geschäftsprozesse in Abhängigkeit von Dokumenteninhalten steuern lassen können. Diese Ansätze sollen dann unter wirklichkeitsnahen Bedingungen evaluiert werden. Bei der Betrachtung der Dokumente soll ihr Inhalt (nicht ihre Erscheinung) als grundlegendes Merkmal der Betrachtung herangezogen werden.

Merkmale der Geschäftsprozesse, denen die Dokumente zugeordnet werden, sollen aus Geschäftsprozessmanagementsystemen (BPM-Systemen) extrahiert werden.

Daraus leiten sich die folgenden Forschungsfragen ab:

1. Wie lassen sich IT-gestützte Geschäftsprozesse durch eingehende Geschäftsdokumente steuern?
2. Wie gut oder schlecht eignen sich hierfür „klassische“ Klassifikationsverfahren?
3. Lassen sich unter Berücksichtigung semantischer Aspekte bessere / zuverlässigere Ergebnisse erzielen?

Um die Forschungsergebnisse im praktischen Gebrauch verwenden zu können, soll eine konkrete Implementierung als Softwaresystem stattfinden. Abbildung 18.1 stellt die Arbeitsweise des Systems schematisiert dar.

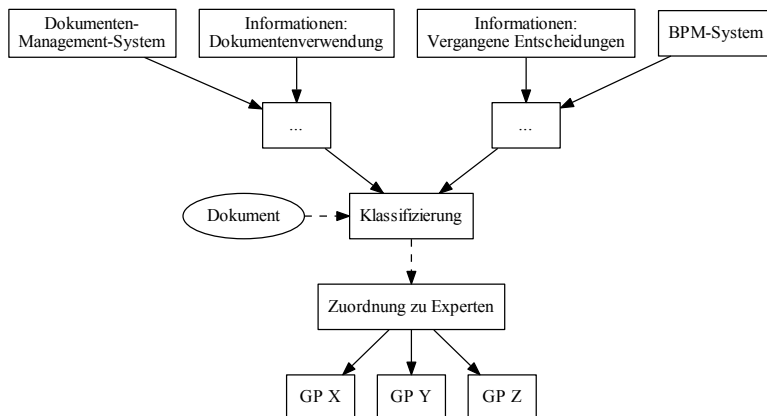


Abbildung 18.1: Funktionsweise des Zielsystems

Aktueller Stand, Verwandte Forschungsbereiche

In der Praxis werden Dokumente im Geschäftsbereich häufig „Fällen“ zugeordnet. Solch ein Vorgehen wird auch als Adaptive Case Management (ACM) [HK11] beschrieben. Der jeweilige Sachbearbeiter bildet mit seiner individuellen Wissensbasis (siehe [VHAA11]) den Knotenpunkt für Informationen zum Fall.

Dokumentenmanagement- und ERP-Systeme bieten häufig auch die Möglichkeit, eine automatische Zuordnung von Dokumenten zu Geschäftsprozessen – oder gar Workflows – vornehmen zu lassen. Sie wenden dabei zwei prinzipielle Verfahren an:

- Bei der Verwendung von *Codierungen* stehen dem externen Sender Informationen zur Verfügung, die die weitere Kommunikation kennzeichnen, z.B. Fallnummern. Dies ist allerdings nicht für die erste Kommunikation mit der Organisation geeignet.
- Bei Verfahren der *musterbasierten Vorlagenerkennung* ist das Layout das Hauptmerkmal der Analyse. Algorithmen aus Bildverarbeitung und Texterkennung extrahieren die relevanten Informationen (z.B. Beträge). Es wird eine Menge von gleich aussehenden Dokumenten benötigt, um die jeweiligen Positionen zu bestimmen. Dadurch ist auch dieses Verfahren nicht für eine erste Kommunikation geeignet.

Im Folgenden soll ein skizzenhafter Überblick über die tangierten Forschungsrichtungen gegeben werden.

Klassifizierungsverfahren

Kern der Forschungsarbeit bilden Klassifizierungsverfahren, wie sie in [Seb02] und [Mit97] beschrieben werden. Sie lassen sich in probabilistische und nichtprobabilistische Verfahren unterteilen.

Probabilistische Verfahren sind unter anderem das Naïve-Bayes-Verfahren, wie es in Spam-Filtern weit verbreitet ist, deutlich komplexere Bayes'sche Netzwerke (einschließlich Hidden Markov Models) und Entscheidungsbäume (C4.5 / ID3).

Nichtprobabilistische Verfahren wie Neuronale Netze oder Support Vector Machines geben demgegenüber keine Zuverlässigkeitsinformation über ihre Zuteilung aus.

Alle diese Verfahren sind in verschiedensten Teilgebieten verbreitet, ihre Verwendung zur Zuordnung von Geschäftsdokumenten zu Prozessen ist jedoch ein neues Forschungsfeld.

Natural Language Processing

Natural Language Processing (NLP) als allgemeines Forschungsfeld beschäftigt sich mit der maschinellen Verarbeitung von Informationen die – wie bei den betrachteten Dokumenten – in natürlicher Sprache vorliegen. Dabei gibt es eine Reihe von Herausforderungen, zu denen geforscht wird.

Ein Hauptproblem stellt die Herstellung von *Eindeutigkeiten* dar, die von Worten auf die dahinterliegenden Konzepte schließen lassen. Mit Hilfe umfangreicher Text-

korpora wird versucht, strukturierte Ontologien aufzubauen [Man04]. Hierzu gehört auch die Erkennung von Synonymen [BHP09] und Eigennamen [TH09], die für die Bedeutung eines Textes ausschlaggebend sind.

Sind solche Abstraktionen der Dokumente vorhanden, bietet das gute Ansatzpunkte für eine weitere Verarbeitung, wie Klassifikation [EA⁺09] oder eine automatisierte Zusammenfassung [SAL10].

Geschäftsprozessanalyse

Um die Verknüpfung zwischen Dokumenten und Geschäftsprozessen herstellen zu können, ist es zweckdienlich, auch die Prozesse in eine einheitliche Darstellungsform zu überführen. Hierzu gibt es die Möglichkeit, schon bei dem Entwurf der Prozesse (Top Down) passende Darstellungen zu wählen ([Lau10]). Solche Vorgehen sind unter Umständen nicht durchsetzbar. Es gibt jedoch auch Ansätze, aus bestehenden Umgebungen Informationen zu gewinnen (z.B. [BY⁺09]).

Diese Extraktion semantischer Aspekte ist insbesondere für die Beantwortung der dritten Forschungsfrage von Interesse.

Literaturverzeichnis

- [BHP09] Bhagat, R; Hovy, E; Patwardhan, S. Acquiring paraphrases from text corpora. In *Proceedings of the fifth international conference on Knowledge capture*, S. 161-168, New York, 2009.
- [BY⁺09] Burstein, M H; Yaman, F, Laddaga, R M; Bobrow, R J. POIROT: acquiring workflows by combining models learned from interpreted traces. In *Proceedings of the fifth international conference on Knowledge capture*, S. 129-136, New York, 2009.
- [EA⁺09] Echarte, F; Astrain, J J; Córdoba, A; Villadangos, J; Labat, A. ACoAR: a method for the automatic classification of annotated resources. In *Proceedings of the fifth international conference on Knowledge capture*, S. 181-182, New York, 2009.
- [HK11] Herrmann, C; Kurz, M. Adaptive Case Management: Supporting Knowledge Intensive Processes with IT Systems. In Schmidt, W (Hrsg.), *S-BPM ONE - Learning by Doing - Doing by Learning. Communications in Computer and Information Science*, 213:80-97, Springer Berlin Heidelberg, 2011.
- [Lau10] Lautenbacher, F. *Semantic Business Process Modeling: Principles, Design Support and Realization (Berichte aus der Informatik)*. Dissertation, Universität of Augsburg, 2010.
- [Man04] Mani, I. Automatically inducing ontologies from corpora. In *Proceedings of CompuTerm 2004: 3rd International Workshop on Computational Terminology*, S. 47-54, 2004.

- [Mit97] Mitchell, T M. *Machine Learning*. McGraw-Hill, Inc., New York, 1997.
- [Seb02] Sebastiani, F. Machine learning in automated text categorization. *ACM Computing Survey*, 34(1):1-47, 2002.
- [SAL10] Shiells, K; Alonso, O; Lee, H J. Generating document summaries from user annotations. In *Proceedings of the third workshop on Exploiting semantic annotations in information retrieval*, S. 25-26, 2010.
- [TH09] Tomanek, K; Hahn, U. Reducing class imbalance during active learning for named entity annotation. In *Proceedings of the fifth international conference on Knowledge capture*, S. 105-112, New York, 2009.
- [VHAA11] Volda, A; Harmon, E; Al-Ani, B. Homebrew databases: complexities of everyday information management in nonprofit organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, S. 915-924, New York, 2011.



Dominik Reichelt, M.Eng, ist wissenschaftlicher Mitarbeiter der Forschungsgruppe Information Management des Instituts für Informationssysteme der Hochschule Hof. Als Software-Engineer war er an verschiedenen Projekten beteiligt. Ihr Spektrum reichte von betrieblichen Reporting-Lösungen bis hin zur Entwicklung von Systemen zur automatisierten Testdurchführung. Neben der Forschung hält er Vorlesungen im Informatikbereich.

Dieser Beitrag ist erschienen in: Thorsten Claus und Niels Seidel (Hrsg.), *Werkstatt europäischen Denkens – 20 Jahre Internationales Hochschulinstitut Zittau*, TUDpress, Dresden, 2014. Online verfügbar: <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-152345>.