

Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache

Timm Caporale¹

Abstract: Organisationen passen ihre Geschäftsprozesse immer öfter und schneller an die sich kontinuierlich ändernden Rahmenbedingungen an. Dabei werden die Prozesse üblicherweise zunächst modelliert, dann analysiert und dann implementiert. Eine bleibende Herausforderung in diesem Prozess betrifft die angemessene Einbeziehung der Wissensträger bzw. Stakeholder. Im Rahmen dieses Beitrags wird eine Modellierungstechnik vorgestellt, die natürliche Sprache als Grundlage der Geschäftsprozessmodellierung verwendet. Die Technik benutzt Natural Language Processing-Techniken, die durch die Verwendung von Satzschablonen aus kontrollierter natürlicher Sprache Geschäftsprozessmodelle erzeugen. Es wird hierzu eine kontextfreie Grammatik entworfen, ein Parser erzeugt, der erzeugte Parsebaum analysiert und als Geschäftsprozessmodell ausgegeben. Mit diesem Ansatz kann garantiert werden, dass die in natürlicher Sprache vorliegende textuelle Beschreibung exakt die Semantik des Geschäftsprozessmodells repräsentiert. Stakeholder, die am Modellierungsprozess beteiligt sind, aber keine Expertise in der Geschäftsprozessmodellierung besitzen, können mit Hilfe der textuellen Beschreibung die Semantik des Geschäftsprozessmodells erfassen.

Keywords: business process modeling, natural language processing

1 Motivation und Problemstellung

Externe Rahmenbedingungen wie beispielsweise der globale Wettbewerbsdruck, komplexe Geschäftsmodelle und der Wunsch nach einer stärkeren Kundenorientierung führen dazu, dass Organisationen ihre Strategie und damit auch ihre Geschäftsprozesse ständig anpassen müssen. Methoden für diese Anpassung bietet das Geschäftsprozessmanagement (BPM). Der natürlichen Sprache kommt dabei in mehreren Phasen eine große Bedeutung zu. In der Identifikationsphase zum Beispiel werden viele Dokumente in natürlicher Sprache erzeugt und analysiert. Auch in den Interviewtechniken, die zur Identifikation der Geschäftsprozesse benutzt werden, ist die natürliche Sprache Grundlage des Verständnisses. Existierende Arbeiten unterstreichen die Relevanz der natürlichen Sprache für das Verständnis von Geschäftsprozessmodellen. Deshalb empfehlen auch viele Arbeiten sich bei der Formulierung textueller Beschriftungen in Geschäftsprozessen an eine bestimmte Syntax zu halten. Viele Wissensträger haben aber leider keine oder nicht ausreichende Erfahrungen im Bereich BPM, um Geschäftsprozessmodelle eigenständig zu erstellen und zu validieren, so dass das Vorgehensmodell der meisten existierenden Ansätze im

¹ Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Kaiserstr. 89, 76133 Karlsruhe, timm.caporale@kit.edu

Allgemeinen vorsieht, dass in einem ersten Schritt die Geschäftsprozessmodellierer das Prozesswissen einer Organisation durch zum Beispiel (un-)strukturierte Experteninterviews oder Workshops von den Wissensträgern extrahieren und daraus Geschäftsprozessmodelle ableiten. Die Validierung inwiefern die textuellen Beschreibungen mit den erzeugten Geschäftsprozessmodellen übereinstimmen, bleibt dabei für die Wissensträger eine große Herausforderung.

Der Forschungsbereich der Spracherkennung (Natural Language Recognition), der in den letzten Jahren große Fortschritte gemacht hat, entwickelt zunehmend neue Techniken, die eine Verarbeitung natürlicher Sprache durch IT-Systeme ermöglichen (Natural Language Processing). Für das Geschäftsprozessmanagement ergeben sich hieraus neue Möglichkeiten. Dieser Beitrag stellt einen Ansatz zur Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache vor. Durch die direkte Verbindung zwischen Geschäftsprozessmodell und natürlichsprachlicher (nicht formaler) Beschreibung sollen Wissensträger besser in die Geschäftsprozessmodellierung einbezogen werden und darüber hinaus die Übereinstimmung zwischen Prozessbeschreibung und Prozessmodell garantiert werden.

Der Beitrag ist wie folgt strukturiert: Abschnitt 2 stellt das Lösungskonzept dar, Abschnitt 3 zählt die wichtigsten verwandten Arbeiten auf und schließt mit einem Vergleich. Abschnitt 4 gibt einen Ausblick.

2 Lösungskonzept

Der Ansatz zur Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache basiert auf der Verwendung eines LL(k)-Parsers für kontextfreie Grammatiken. Der Parser und seine nachgelagerten Analysekomponenten sind für die Verarbeitung der natürlichen Sprache und die Transformation in ein Geschäftsprozessmodell verantwortlich. Der im Rahmen des Ansatzes verwendete Parser wurde mit Hilfe von ANTLR² als LL(*)-Parsergenerator erstellt. ANTLR erwartet eine kontextfreie Grammatik in angepasster erweiterter Backus-Naur-Form. Anforderungen zur Erstellung dieser kontextfreien Grammatik waren, dass alle einfachen Kontrollflussmuster der Workflowpatterns³ integriert werden und der Ausdrucksraum erweiterbar bleiben sollte. Für die englische Sprache wurde eine entsprechende Grammatik entwickelt, die Bedingungen, Aktivitäten und den Kontrollfluss (Sequenz, Parallel Split, Synchronization, Exclusive Choice, Simple Merge) darstellt. Die Analysekomponenten des Parsers erstellen dann auf Basis des aktuellen Parserkontexts eine Vorhersage für das nächste (Nicht-) Terminalsymbol und zeigen es dem Benutzer in Form einer Auto-Vervollständigungsfunktion und als dynamische Satzschablone an.

² <http://www.antlr.org/>

³ <http://www.workflowpatterns.com/patterns/control/>

Satzschablonen dienen dazu, die natürliche Sprache einzuschränken und Formulierungen kontrolliert und bewusst zu verwenden. Mit kontrollierter natürlicher Sprache ist in diesem Zusammenhang die Einschränkung gemeint, dass nur Eingaben akzeptiert werden, die auf den erzeugten Satzschablonen basieren und somit vom Parser auch ausgewertet werden können. (Statische) Satzschablonen werden beispielsweise im Requirements Engineering verwendet, um Softwarefunktionen zu spezifizieren. Während im Requirements Engineering jede Anforderung für sich steht und diese idealerweise auch vollständig beschreibt, ist jeder Teil einer textuellen Beschreibung eines Geschäftsprozesses für sich kontextsensitiv. Die Satzschablonen, die im Rahmen der Modellierung verwendet werden, müssen daher zur Laufzeit dynamisch an den aktuellen Kontext angepasst werden. Parser und Satzschablonen basieren auf folgender Grammatik in eBNF:

```
program: sentence*;
sentence: placestart | transstart;
placestart: plstartlist placeslist ', then I do '
           transitionlist '. ';
plstartlist: 'If ' | 'As soon as ';
placeslist: place | eitherpl place orpl furtherplaces |
           place andpl furtherplaces;
furtherplaces: place | place orpl furtherplaces | place
              andpl furtherplaces;
eitherpl: 'either ';
orpl: ' or ';
andpl: ' and ';
transitionlist: transition | eithertr transition ortr
              furthertransit | transition andtr furthertransit;
furthertransit: transition | transition ortr
              furthertransit | transition andtr furthertransit;
eithertr: 'either ';
ortr: ' or ';
andtr: ' and ';
transstart: lasttrans transfollow '. ' | 'After I did '
           transitionlist ', ' transfollow '. ';
lasttrans: 'After that, ';
transfollow: plfollowlist placeslist | trfollowlist
            transitionlist;
plfollowlist: 'I have ' | 'I get ';
trfollowlist: 'I do ';
place: content;
transition: content;
content: STRING+ ( ' ' | STRING )* | content '.' content;
```

Neben der Vorhersage des nächsten (Nicht-)Terminalsymbols wird im Rahmen der entwickelten Methode der bisher eingegebene Text ausgewertet und in ein einfaches BPMN-Modell transformiert. Abbildung 1 zeigt das Parserkonzept basierend auf der kontextfreien Grammatik und des eingegebenen Textes in kontrollierter natürlicher Sprache. Zur Transformation der natürlichen Sprache in ein grafisches Prozessmodell wird von der Analysekomponente ein Graphviz⁴-konformer Text erstellt. Dieses dient der Komponente als Grundlage zur Erzeugung der Geschäftsprozessmodellgrafik.

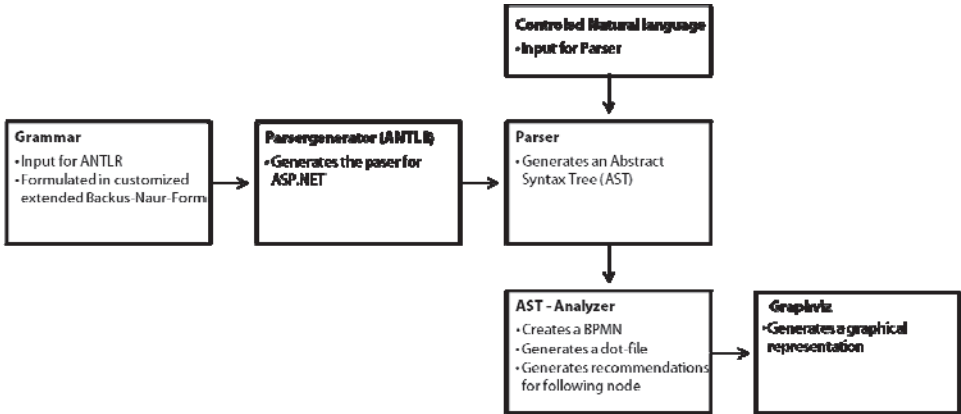


Abb. 1: Transformationskonzept

Zur Veranschaulichung der Funktionalität wurde ein Prototyp als Webapplikation entwickelt⁵ und erste Tests durchgeführt. Die Tests unter Laborbedingungen zeigen eine korrekte Auswertung und Transformation [Ca16]. Die Entwicklung des Prototyps als Webapplikation ermöglicht außerdem, Anwender schnell und einfach in die Modellierungsaktivitäten einzubinden.

3 Verwandte Arbeiten

Die existierenden verwandten Ansätze lassen sich in zwei Kategorien unterteilen: Auf der einen Seite existieren Ansätze, die natürliche Sprache vor oder bei der Modellierung analysieren. Dabei sollen Informationen der natürlichen Sprache verarbeitet werden und bei der Modellierung Berücksichtigung finden. Auf der anderen Seite existieren Ansätze, die die natürliche Sprache in den Modellen analysieren. Dabei handelt es sich meistens um eine Analyse der textuellen Beschriftungen. Ansätze der ersten Kategorie verwenden zumeist NLP-Techniken um aus bereits existierenden Beschreibungen Geschäftsprozessmodelle, ER-Modelle oder UML-Diagramme zu generieren. Alle Arbeiten haben gemeinsam, dass sie Standard-NLP-Tools wie WordNet, Tagger, Parser

⁴ <http://www.graphviz.org>

⁵ <https://bpm.caporale.eu/>

oder existierende Spracherkennungsverfahren verwenden. Für die Erzeugung von Geschäftsprozessmodellen existieren z.B. Ansätze, die mit Hilfe von Parsern Texte analysieren und daraus (halb-) automatisch BPMN Modelle ableiten z.B. [FMP11, GKC07, GSB09, SP10]. Ähnliche Ansätze existieren für Abhängigkeitsdiagramme [Ga01], Entity-Relationship-Modelle [OHA08], UML-Diagramme [DB09, MP12] und Petri-Netze [Ke14]. Einige Autoren beschäftigen sich auch mit abstrakteren Ansätzen, die unabhängig von der gewählten Modellierungssprache sind [FKM07]. Ansätze der zweiten Kategorie analysieren die Beschriftungen auf z.B. terminologische Fragestellungen, um die Gesamtqualität der Modelle zu verbessern und basieren zumeist auf WordNet [BPB09, KHO11]. Einige proprietäre Ansätze gehen über die Analyse syntaktischer Aspekte hinaus und suchen nach semantischen Fehlern [GL11] oder erstellen semantische Annotationen [BSP08]. Um eine Verbindung zwischen Modellen und der natürlichen Sprache herzustellen, realisieren weitere Ansätze die Erzeugung von natürlich sprachlichem Text für UML-Klassendiagramme [MAA08], Objektmodelle [LRR96], Geschäftsprozessmodelle [Ke14] und konzeptuelle Modelle [Da92].

Der im Rahmen dieses Beitrags vorgestellte Ansatz vermischt beide Kategorien. Auf der einen Seite wird ein Parser verwendet, der aus natürlicher Sprache ein Prozessmodell erzeugt (ähnlich zu [Ke14]). Auf der anderen Seite wird die Qualität des erstellten Prozessmodells durch die Verknüpfung zur textuellen Beschreibung sichergestellt. Im Gegensatz zu den meisten existierenden Ansätzen benutzt der Parser keine stochastischen Methoden, um die Semantik des Textes zu erfassen, sondern garantiert durch den festgelegten Eingaberaum (in Form der kontrollierten natürlichen Sprache) eine fehlerfreie und verlustfreie Transformation. Ein Nachteil ist dabei, dass existierende Beschreibungen durch die vorgestellte Technik nicht geparkt werden können.

4 Ausblick

Im Rahmen dieses Beitrags wurde eine Technik vorgestellt, welche natürliche Sprache als Eingabe zur Modellierung von Geschäftsprozessen verwendet. Die so erstellte textuelle Beschreibung repräsentiert exakt die Semantik des Geschäftsprozessmodells. Realisiert wurde die Technik durch den Entwurf einer kontextfreien Grammatik als Parsergrundlage und die Verwendung von Satzschablonen zur Kontrolle der natürlichen Sprache. Der Prototyp zeigt unter Laborbedingungen vielversprechende Resultate.

Die nächsten Schritte im Rahmen der Forschungsarbeit betreffen vor allem die Anpassung der Satzschablonen um weitere und flexiblere Satzstrukturen. Durch den Einsatz moderner Anwendungsoberflächen und die Integration weiterer benutzerunterstützender Komponenten könnte außerdem untersucht werden, inwiefern IT-affine Stakeholder eigenständig Geschäftsprozessmodelle erstellen können. Als Ausblick ist denkbar, dass durch die textuelle Eingabe neue Anwendungsfelder entstehen. Eine Anpassung der Technik zur Modellierung von Anforderungen für Softwaresysteme oder Richtlinien können Anwendungsfälle außerhalb der BPM-Domäne sein.

Literaturverzeichnis

- [BPB09] Breuker, D.; Pfeiffer, D.; Becker, J.: Reducing the Variation in Intra-and Interorganizational Business Process Modeling-An Empirical Evaluation, 2009.
- [BSP08] Bögl, A. et al.: Semantic annotation of epc models in engineering domains to facilitate an automated identification of common modelling practices: Enterprise Information Systems. Springer, 2008; S. 155–171.
- [Ca16] Caporale, T.: A tool for natural language oriented Business Process Modeling. ZEUS, 2016.
- [Da92] Dalianis, H.: A method for validating a conceptual model by natural language discourse generation. Springer, 1992.
- [DB09] Deeptimahanti, D. K.; Babar, M. A.: An automated tool for generating UML models from natural language requirements. IEEE Computer Society, 2009.
- [FKM07] Fliedl, G. et al.: Deriving static and dynamic concepts from software requirements using sophisticated tagging. In Data & Knowledge Engineering, 2007, 61; S. 433–448.
- [FMP11] Friedrich, F.; Mendling, J.; Puhlmann, F.: Process model generation from natural language text. Springer, 2011.
- [Ga01] Gangopadhyay, A.: Conceptual modeling from natural language functional specifications. In Artificial Intelligence in Engineering, 2001, 15; S. 207–218.
- [GKC07] Ghose, A.; Koliadis, G.; Chueng, A.: Process discovery from model and text artefacts. IEEE, 2007.
- [GL11] Gruhn, V.; Laue, R.: Detecting Common Errors in Event-Driven Process Chains by Label Analysis. In EMISA, 2011, 6; S. 3–15.
- [GSB09] Gonçalves, João Carlos de AR; Santoro, F. M.; Baiao, F. A.: Business process mining from group stories. IEEE, 2009.
- [Ke14] Keuter, B.: Bidirektionale Abbildung zwischen Geschäftsprozessmodellen und IT-Kommunikationssystemen. KIT Scientific Publishing, 2014.
- [KHO11] Koschmider, A.; Hornung, T.; Oberweis, A.: Recommendation-based editor for business process modeling. In Data & Knowledge Engineering, 2011, 70; S. 483–503.
- [MAA08] Meziane, F.; Athanasakis, N.; Ananiadou, S.: Generating Natural Language specifications from UML class diagrams. In Requirements Engineering, 2008, 13; S. 1–18.
- [MP12] More, P.; Phalnikar, R.: Generating UML Diagrams from Natural Language Specifications. In International Journal of Applied Information Systems (IJ AIS), Foundation of Computer Science FCS, New York, USA, 2012, 1.
- [OHA08] Omar, N.; Hassan, R.; Arshad, H.; Sahran, S.: Automation of database design through semantic analysis, 2008.
- [SP10] Sinha, A.; Paradkar, A.: Use cases to process specifications in business process modeling notation. IEEE, 2010.