

Integration von Prozessmodellen im Großen: Konzept, Methode und experimentelle Anwendungen

Peter Fettke

Institut für Wirtschaftsinformatik im Deutschen Forschungszentrum für Künstliche Intelligenz
(DFKI GmbH) und Universität des Saarlandes, Saarbrücken, Germany
peter.fettke@iwi.dfki.de

Abstract. Anwender mischen (englisch *merge*) und integrieren digitale Daten routinemäßig mithilfe von Computern. Auch Modellierer haben vielfältigen Bedarf an dieser Grundoperation, beispielsweise im Rahmen der Konsolidierung von Prozessvarianten, der verteilten Modellierung oder der Fusion mehrerer Unternehmen. Auch wenn bereits verschiedene Methoden zur Modellintegration bekannt sind, konzentrieren sich vorherrschende Ansätze auf die Integration *weniger* Prozessmodelle (Integration im Kleinen). Dagegen entstehen bei der Integration *vieler* Prozessmodelle Herausforderungen (Integration im Großen), die bisher nur unzureichend adressiert werden. Der vorliegende Beitrag präsentiert ein Konzept zur Bewältigung dieser Herausforderungen. Das im Kern auf einer hierarchisch-agglomerativen Cluster-Analyse beruhende Konzept wird prototypisch implementiert und anhand ausgewählter Szenarien experimentell angewendet. Die Ergebnisse verdeutlichen, dass die Integration vieler Prozessmodelle durch eine Cluster-Analyse erst praktikabel wird.

Keywords: Verständlichkeit, Komplexität, Kompressionseffekt, Expansionseffekt, Explosionseffekt

1 Ausgangssituation und Problemstellung

Das Mischen (englisch *merge*) und Integrieren digitaler Daten ist eine Grundoperation, an der seit der Verfügbarkeit von Computern ein Bedarf besteht. Während das maschinelle Verarbeiten von Text seit den Ursprüngen der Informatik erforscht wird, ist das Mischen und Integrieren von Prozessmodellen indes ein Arbeitsbereich, der erst seit der allgemeinen Verfügbarkeit von Software-Werkzeugen für die Prozessmodellierung eine Rolle spielt [10].

Dabei kann die Integration von Prozessmodellen in verschiedenen Anwendungskontexten von Nutzen sein; zu nennen ist beispielsweise [14]: (a) die Modellkonsolidierung in verteilten und dezentralen Modellierungsprojekten, (b) das Zusammenführen verschiedener Varianten eines Prozesses innerhalb einer oder mehrerer Organisationen (beispielsweise bei Unternehmensfusionen), (c) das Vereinheitlichen von Perspektiven im Rahmen der multiperspektivischen Prozessmodellierung oder (d) die induktive Erstellung von Referenzprozessen.

Die in den spezifischen Anwendungskontexten erzielten Nutzenaspekte sind vielfältig [14]: Zunächst entstehen Vorteile innerhalb der Modellwelt; mit der Modellintegration geht eine Modellverkürzung und damit Komplexitätssenkung einher. Das Zusammenfügen führt zu einer Harmonisierung der Modellstruktur. Die Vorteile erstrecken sich aber nicht nur auf die Modellwelt, sondern auch auf die Realwelt. Weiterhin profitiert die Gestaltung der Organisation und des IT-Systems. Beispielsweise ermöglicht die Integration im Rahmen der IT-Systemgestaltung die Mehrfachverwendung von DV- und Implementierungskonzepten. Im Rahmen der Organisationsgestaltung verdeutlicht die Integration, wo Prozessverantwortlichkeiten mehrfach vergeben und organisatorische Regeln mehrfach genutzt werden können. Darüber hinaus zeigt die Integration Potentiale auf, wie die in den Modellen beschriebenen Tätigkeiten und Handlungen auch real integriert werden können.

Im Folgenden wird zwischen der Integration *im Kleinen* und der Integration *im Großen* unterschieden. Bei der Integration im Kleinen sind nur zwei oder wenige Prozessmodelle zu integrieren. Bei der Integration im Großen wird von dieser Annahme abgesehen, wodurch neue Fragen entstehen: Ist es sinnvoll, sämtliche Prozessmodelle vollständig zu integrieren? In welcher Reihenfolge sind die Prozessmodelle zu integrieren? Können innerhalb der Prozessmodelle ähnliche Gruppen ausgemacht werden?

Bekannte Methoden der Integration von Prozessmodellen unterstützen den Modellierer nur unzureichend bei den aufgeworfenen Fragen (vergleiche Abschnitt 3). Daher ist es erklärtes Ziel des vorliegenden Beitrags, eine Methode zu entwickeln, welche dem Modellierer auch bei der Integration *vieler* Prozessmodelle eine sinnvolle Unterstützung anbietet. Grundgedanke des Ansatzes ist es, die aufgeworfenen Fragen mithilfe einer Cluster-Analyse zu beantworten. Dabei soll die innovative Methode nicht nur theoretisch konzipiert, sondern auch technisch implementiert und praktisch erprobt werden.

Dieser Beitrag ist das Ergebnis eines gestaltungsorientierten Forschungsansatzes. Im Rahmen des Forschungsprozesses sind mehrere Iterationen durchlaufen worden, um die Methode zu konzipieren, prototypisch zu implementieren, zu testen und praktisch zu erproben. In der vorliegenden Fassung wird der aktuell erreichte konzeptionelle, methodische und technische Stand beschrieben sowie von experimentellen Anwendungen berichtet, welche sowohl die Anwendbarkeit als auch die Nützlichkeit der entwickelten Methode nachweisen.

Der Beitrag ist wie folgt aufgebaut: Nach dieser Einleitung wird in *Kapitel 2* ein Fallbeispiel angeführt, anhand dem die zuvor skizzierten Herausforderungen näher illustriert werden. *Kapitel 3* gibt einen Überblick über den erreichten Forschungsstand. *Kapitel 4*, der Hauptteil der Arbeit, führt die entwickelte Methode ein. Ein Einblick in die Implementierung des Ansatzes liefert *Kapitel 5*. Die experimentelle Anwendung des Ansatzes beschreibt *Kapitel 6*. Abgeschlossen wird der Beitrag mit einer Diskussion der Ergebnisse und einem Resümee in *Kapitel 7*.

2 Fallbeispiel

Ausgangspunkt des Fallbeispiels ist eine Bibliothek, die 80 Modelle realer Verwaltungsabläufe inklusive diverser Prozessvarianten umfasst [17]. Die Größe der Modelle schwankt zwischen 19 und 198 Knoten (Durchschnitt: 79 Knoten). Im Folgenden werden ausgewählte Prozessmodelle mithilfe des von La Rosa et al. [10] beschriebenen Verfahrens integriert. In den Fällen (I) und (II) werden zwei beziehungsweise zehn Varianten eines Verwaltungsprozesses integriert; Fall (III) betrifft die Integration der gesamten Prozessbibliothek.

Tabelle 1 beschreibt die in den drei Integrationsfällen erzielten Ergebnisse, wobei jeweils folgende Angaben gemacht werden: (a) Anzahl der zu integrierenden Ausgangsmodelle; (b) Größe der Ausgangsmodelle, gemessen als Summe der Knoten aller Ausgangsmodelle; (c) durchschnittliche Größe der Ausgangsmodelle, gemessen in Knoten; (d) Größe des integrierten Modells, gemessen in Knoten; (e) Kompressionsrate und (f) Expansionsrate. Die Kompressionsrate beschreibt, inwieweit die Größe des integrierten Modells im Verhältnis zur *Summe* der Größen der Ausgangsmodelle variiert (definiert als Quotient aus (d) und (b)). Die Expansionsrate beschreibt, inwieweit die Größe des integrierten Modells im Verhältnis zur *durchschnittlichen* Größe der Ausgangsmodelle variiert (definiert als Quotient aus (d) und (c)).

Die gemessenen Werte verdeutlichen: (I) Erwartungsgemäß führt das Integrieren zweier Prozessmodelle zu einer Modellkompression bei einem moderaten Anstieg der Expansionsrate. (II) Die Integration von zehn Varianten eines Prozesses erlaubt weiterhin eine umfassende Modellverkürzung. Allerdings hat sich die Größe des Ausgangsmodells weit mehr als versechsfacht. (III) Die Integration aller Prozessmodelle führt zu einer Expansion des integrierten Prozessmodells um den Faktor 40. Gleichwohl kann weiterhin eine Modellkompression von rund 50 Prozent erzielt werden.

Die Befunde verdeutlichen, dass die Modellintegration zwar einerseits zu einer Modellverkürzung führt. Andererseits steigt die Komplexität des integrierten Modells gemessen als die Anzahl der Knoten rasant an. Eine Ursache für eine hohe Expansionsrate liegt darin, dass die Ausgangsmodelle sehr verschieden sind. Folglich ist die Sinnhaftigkeit einer vollständigen Modellintegration zu prüfen, um eine letztlich unsinnige Integration verschiedener Modelle zu vermeiden. Fraglich bleibt allerdings, welche der 80 Ausgangsmodelle in welchem Umfang integriert werden sollen. Ohne eine Unterstützung kann der Modellierer letztlich nur auf Basis intuitiver Vermutungen nach einem *Trial-and-Error*-Verfahren unterschiedliche Integrationsvarianten probeweise durchspielen.

Tabelle 1. Auswirkungen der Integration von Prozessmodellen

Fall	(a) Anzahl der Ausgangsmodelle	(b) Größe der Ausgangsmodelle (Summe)	(c) Größe der Ausgangsmodelle (Durchschnitt)	(d) Größe des integrierten Modells	(e) Kompressionsrate	(f) Expansionsrate
(I)	2	270	135	166	61 %	123 %
(II)	10	1.394	139	916	66 %	657 %
(III)	80	6.327	79	3.154	50 %	3.988 %

3 Stand der Forschung

Die Forschung zur Integration von Modellen hat beginnend mit Arbeiten zur Integration von Datenmodellen eine lange Tradition [3]. Heute existieren verschiedene Forschungsrichtungen, die sich in der Regel an dem jeweils zu integrierenden Modelltyp orientieren. So gibt es beispielsweise Arbeiten zur Integration von Statecharts [12], zu Workflow-Modellen [16] und zu Sequenz-Diagrammen [18].

Eine der ersten Arbeiten, welche die Integration von Prozessmodellen umfassend behandelt, ist die 1996 vorgelegte Dissertation von Rosemann [14]. Rosemann identifiziert (a) die Modellauswahl, (b) die Konfliktbehandlung, (c) die Integration im engeren Sinne („merge“) und (d) die Modellmodifikation als Aufgabenbereiche der Modellintegration. Zwar liefert Rosemann im Rahmen der Modellauswahl verschiedene Ideen zur Integration mehrerer Prozessmodelle, allerdings bleiben diese noch auf einer verhältnismäßig allgemeinen Ebene.

In den nachfolgenden Jahren wurden unterschiedliche Ansätze zur Konkretisierung der von Rosemann identifizierten Aufgabenbereiche unterbreitet, wobei sich bisher keine einheitliche Terminologie in diesem Zusammenhang herausgebildet hat. Beispielsweise subsumieren manche Autoren unter dem Mischen von Modellen auch Aufgaben zur Konfliktbehandlung.

Tabelle 2 stellt überblicksartig die Schwerpunkte ausgewählter Arbeiten zur Integration von Prozessmodellen anhand der vier von Rosemann aufgezeigten Teilbereiche der Integration von Prozessmodellen dar. Die Übersicht verdeutlicht, dass die Modellauswahl zwar bereits von Rosemann 1996 thematisiert wurde. Gleichwohl konzentrieren sich die nachfolgenden Arbeiten zur Modellintegration schwerpunktmäßig auf die Konfliktbehandlung, die Integration im engeren Sinne und auf die Modellmodifikation. Diese Forschungslücke wird exemplarisch an der aktuellen Arbeit von La Rosa et al. [10] näher erläutert.

La Rosa et al. stellen ein neues Verfahren zur Integration von Prozessmodellen vor, das insbesondere folgenden Anforderungen genügt: (1) Das Verhalten der Ausgangsmodelle ist vollständig im integrierten Modell enthalten. (2) Die Ausgangsmodelle können aus dem integrierten Modell rekonstruiert werden. Die Autoren stellen nicht nur ein Konzept vor, sondern bieten auch eine lauffähige Implementierung an. Um mehrere Prozessmodelle zu integrieren, schlagen die Autoren die mehrmalige Integration zweier

Tabelle 2. Schwerpunkte vorhandener Arbeiten zur Integration von Prozessmodellen

Autor / Quelle	Jahr	Modellauswahl	Konfliktbehandlung	Integration im engeren Sinne	Modellmodifikation
Rosemann [14]	1996	⊙	●	●	⊙
Mendling, Simon [11]	2006	○	⊙	●	⊙
Gottschalk et al. [9]	2008	○	⊙	●	⊙
Pfeiffer [13]	2008	○	●	○	○
La Rosa et al. [10]	2013	○	●	●	⊙

Legende: ○ wird nicht oder unwesentlich behandelt, ⊙ wird ansatzweise behandelt, ● wird umfassend behandelt

Prozessmodelle vor. Ob eine derartige Integration überhaupt sinnvoll ist oder in welcher Reihenfolge die Prozessmodelle zu integrieren sind, wird von den Autoren nicht thematisiert. Vor diesem Hintergrund gilt das Verfahren von La Rosa et al. zwar als leistungsfähig, bietet allerdings letztlich keine Lösung für die aufgezeigten Herausforderungen. Auch die weiteren in Tabelle 2 genannten Verfahren betrachten die Modellauswahl wenig bis gar nicht.

Das im vorliegenden Beitrag vorgestellte Konzept basiert auf Techniken der Cluster-Analyse, die bereits in anderen Kontexten der Prozessmodellierung eingesetzt werden [1, 15, 17]. Diese Ansätze inspirieren einerseits aufgrund eines ähnlichen Lösungswegs die vorliegende Arbeit. Andererseits verfolgen sie andere Zwecke als eine Modellintegration und sind daher für die vorliegende Arbeit nicht von direkter Bedeutung.

Zusammenfassend kann festgehalten werden, dass zwar unterschiedliche Arbeiten zur Integration und Cluster-Analyse von Prozessmodellen existieren. Allerdings bleiben die skizzierten Herausforderungen der Integration *vieler* Prozessmodelle bei den bisherigen Arbeiten weitgehend unberücksichtigt. Folglich besteht eine Forschungslücke, die mit der vorliegenden Arbeit geschlossen wird.

4 Konzeption des Ansatzes

4.1 Theoretische Vorüberlegungen

Die in Abschnitt 2 anhand eines Fallbeispiels illustrierten Wirkungen der Integration von Prozessmodellen werden im Folgenden verallgemeinert. Das Fallbeispiel hat verdeutlicht, dass zwei gegenläufige Entwicklungen bei der Integration von Prozessmodellen zu beachten sind. Erstens entsteht ein Kompressionseffekt, das heißt, die Größe des integrierten Modells ist kleiner als die Summe der Größen der Ausgangsmodelle. Voraussetzung für den Kompressionseffekt ist, dass zwischen den Ausgangsmodellen gewisse Ähnlichkeiten und Redundanzen bestehen. Diese Redundanzen werden im integrierten Modell vermieden, sodass eine kleinere Modellgröße resultiert. Der Kompressionseffekt tritt nicht ein, wenn die Ausgangsmodelle vollständig verschieden sind. Wenn dagegen die Ausgangsmodelle in jeder Hinsicht gleich sind, entsteht ein maximaler Kompressionseffekt.

Der Kompressionseffekt ist im Rahmen der Modellintegration wünschenswert. Allerdings entsteht darüber hinaus ein *Expansionseffekt*, das heißt, die Größe des integrierten Modells ist größer als die durchschnittliche Größe der Ausgangsmodelle. Dieser Effekt entsteht dadurch, dass die Ausgangsmodelle auch Unterschiede enthalten, die in ihrer Gesamtheit in das zu integrierende Modell aufzunehmen sind und damit dieses vergrößern. Der Expansionseffekt tritt nicht ein, wenn die Ausgangsmodelle in jeder Hinsicht gleich sind. Wenn dagegen die Ausgangsmodelle keine Gemeinsamkeiten enthalten, entsteht ein maximaler Expansionseffekt. Der Expansionseffekt ist im Gegensatz zum Kompressionseffekt *nicht* wünschenswert.

Kompressions- und Expansionseffekte verhalten sich gegenläufig, sodass der Modellierer beide Effekte im Rahmen der Integration auszubalancieren hat. Inwiefern ein bestimmter Kompressions- und Expansionseffekt akzeptabel oder wünschenswert ist,

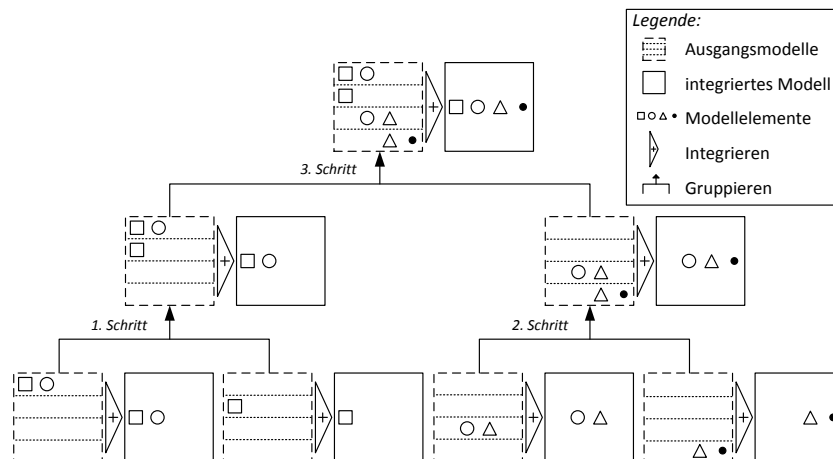


Abbildung 1. Grundidee der vorgeschlagenen Methode

lässt sich kaum allgemeingültig beantworten. Beispielsweise können bei der Beurteilung eines sinnvollen Integrationsgrades auch weitere Faktoren wie organisatorische Verantwortlichkeiten und Rahmenbedingungen eine bedeutende Rolle spielen. Folglich ist eine abschließende Bewertung letztlich nur im Einzelfall möglich.

Indes entsteht dabei ein weiteres Problem: Die Anzahl möglicher Gruppierungen aller Ausgangsmodelle entspricht dem kombinatorischen Problem der Bestimmung aller Partitionen einer Menge. Mit zunehmender Mächtigkeit einer Menge steigt die Anzahl möglicher Partitionen exponentiell an (*Explosionseffekt*). So gibt es beispielsweise bereits 52 Integrationsvarianten bei fünf Ausgangsmodellen, über 100.000 Varianten bei zehn Ausgangsmodellen und ab zwölf Ausgangsmodellen bereits mehrere Millionen Integrationsvarianten (siehe <http://oeis.org/A000110>). Damit scheidet nicht nur ein manuelles Durchmuster aller möglicher Gruppierungen und einer probeweisen Integration sowie Beurteilung aus. Vielmehr sind auch einer maschinellen Analyse des exponentiell ansteigenden Lösungsraums praktische Grenzen gesetzt. Das im Folgenden beschriebene Verfahren bietet einen praktikablen Weg, um den Kompressions- und Expansions- sowie Explosionseffekt geeignet zu beherrschen.

4.2 Grundidee der Methode und Überblick

Abbildung 1 visualisiert die Grundidee der Methode zur Integration einer Menge von Prozessmodellen. Die Integration wird schrittweise vorgenommen. Im ersten Schritt liegen die zu integrierenden Ausgangsmodelle vor, die trivialerweise jeweils auch ein teilweise integriertes Modell bilden. In jedem weiteren Schritt ist zu entscheiden, welche Modelle weiter zu integrieren sind. Dabei werden jeweils diejenigen teilweise integrierten Modelle ausgewählt, deren Integration als besonders günstig zu beurteilen

ist, also vermutlich eine möglichst hohe Kompressions- und niedrige Expansionsrate aufweisen. Diese Schritte werden wiederholt, bis sämtliche teilweise integrierten Modelle vollständig integriert sind. Anschließend ist zu entscheiden, welche teilweise integrierten Modelle im konkreten Anwendungsfall akzeptabel sind.

Der geschilderte Integrationsprozess ist analog dem Vorgehen einer hierarchisch-agglomerativen Cluster-Analyse [8], wobei die zu integrierenden Ausgangsmodelle als Cluster-Objekte gewählt werden. Bei dieser Art der Cluster-Analyse bilden zunächst sämtliche Cluster-Objekte einen eigenständigen Cluster. In den folgenden Schritten werden jeweils diejenigen Cluster fusioniert, die eine hohe Ähnlichkeit zueinander besitzen. Die Cluster-Fusionierung ist abgeschlossen, wenn sämtliche Cluster zu einem einzigen Cluster fusioniert sind. Damit fügt sich dieser Typ der Cluster-Analyse konzeptionell in den zuvor geschilderten Integrationsprozess ein.

Im Folgenden wird die Grundidee der Methode weiter konkretisiert und detailliert beschrieben. Dabei werden fünf Schritte unterschieden: In *Schritt 1* sind zunächst die Ausgangsmodelle einheitlich aufzubereiten (Abschnitt 4.3). Zwischen den Ausgangsmodellen wird die Ähnlichkeit in *Schritt 2* bestimmt (Abschnitt 4.4). Auf Basis der Ähnlichkeiten wird eine hierarchisch-agglomerative Cluster-Analyse in *Schritt 3* durchgeführt (Abschnitt 4.5). Die Cluster-Struktur dient in *Schritt 4* zur schrittweisen Integration der Ausgangsmodelle (Abschnitt 4.6). Mithilfe ausgewählter Kennzahlen werden die Ergebnisse in *Schritt 5* interpretiert und ein sinnvoller Integrationsgrad festgelegt (Abschnitt 4.7).

4.3 Schritt 1: Aufbereitung der Ausgangsmodelle

Das *Ziel* des ersten Schrittes besteht in der Vorbereitung des Integrationsprozesses und der Aufbereitung der Ausgangsmodelle. Im Wesentlichen ist dafür zu sorgen, dass die Prozessmodelle in einem einheitlichen Ausgangsformat vorliegen. Ferner ist es sinnvoll, syntaktische und semantische Mängel in den Ausgangsmodellen zu beheben. Typische Probleme sind beispielsweise Rechtschreibfehler in den Bezeichnern der Aktivitäten, fehlende oder überflüssige Kanten zwischen Modellelementen oder falsch gewählte Konnektoren. Ohne derartige Korrekturen kann sich die Qualität der weiteren Ergebnisse verschlechtern.

Diese vorbereitenden Schritte können manuell oder (teil-)automatisiert durchgeführt werden. Beispielsweise existieren Vorschläge zur Transformation verschiedener Modellierungssprachen und zur Prüfung von Modellen [5, 6].

Im Folgenden wird davon ausgegangen, dass das gewählte Format zur Repräsentation von Prozessmodellen mindestens folgende Informationen umfasst: Funktionen, Ereignisse und Konnektoren sowie den notwendigen Kontrollfluss zwischen diesen Elementen.

Ergebnis ist eine Menge von Prozessmodellen in einem einheitlichen Format.

4.4 Schritt 2: Bestimmung der Ähnlichkeit

Das *Ziel* des zweiten Schrittes besteht in der Quantifizierung der Ähnlichkeit der Ausgangsmodelle. Eine solche Ähnlichkeit wird im Rahmen der Cluster-Analyse benötigt,

um jeweils ähnliche Objekte gruppieren zu können. Im Rahmen der Prozessmodellierung werden derartige Ähnlichkeits- oder auch Distanzmaße typischerweise nicht mit erhoben, sodass ein bestimmtes Vorgehen zu wählen ist.

Grundsätzlich kann die Ähnlichkeit zwischen Prozessmodellen hinsichtlich unterschiedlicher Aspekte ermittelt werden [4]. Prozessmodelle können in einer Hinsicht ähnlich und in einer anderen Hinsicht unähnlich sein. Im Hinblick auf die verfolgte Zielsetzung der Integration von Prozessmodellen spielen insbesondere zwei Aspekte eine Rolle:

- Ähnlichkeit einzelner Prozessschritte: Ein Prozessmodell beschreibt verschiedene Prozessschritte. Die in zwei unterschiedlichen Prozessmodellen beschriebenen Prozessschritte können ähnlich oder unterschiedlich sein.
- Ähnlichkeit der methodischen Ordnung: Ein Prozessmodell definiert eine bestimmte zeitlich-sachlogische Ordnung zwischen den im Prozessmodell beschriebenen Prozessschritten. Diese Ordnung kann ähnlich, aber auch deutlich verschieden sein.

Die Ähnlichkeit einzelner Prozessschritte hat bei der Integration von Prozessmodellen eine hohe Bedeutung. Prozessmodelle, die keine oder nur wenige ähnliche Prozessschritte aufweisen, sind für eine Integration von Prozessmodellen von nur geringer Bedeutung. Die Bestimmung dieser Ähnlichkeit kann trivial sein, wenn die Korrespondenzen zwischen Modellelementen bereits bei der Prozessmodellierung festgelegt wurden. Das ist häufig dann der Fall, wenn ein kontrolliertes Vokabular verwendet wird. Andernfalls ist eine Identifikation von Korrespondenzen über manuelle oder maschinelle Verfahren herzustellen [5].

Die Ähnlichkeit der methodischen Ordnung kann unterschiedlich bestimmt werden. Im Wesentlichen lassen sich vier Gruppen unterscheiden [4]: (1) Graphen-basierte Verfahren, (2) Abhängigkeiten zwischen einzelnen Prozessschritten, (3) Ähnlichkeiten von theoretisch möglichen oder (4) tatsächlich beobachteten Prozessabläufen. Im Hinblick auf die Integration von Prozessmodellen ist grundsätzlich eine ähnliche methodische Ordnung erstrebenswert, da dies bei der Integration der Prozessmodelle zu entsprechenden Integrationspotentialen führt. Dabei gilt es zu berücksichtigen, dass eine Graphen-basierte Ähnlichkeit weder notwendig noch hinreichend für eine ähnliche methodische Ordnung von Prozessschritten ist [2]. Die exakte Bestimmung der theoretisch möglichen Abläufe ist zuweilen algorithmisch aufwändig, da die Menge der möglichen Abläufe unendlich werden kann. Andererseits liegen tatsächlich beobachtete Prozessabläufe nicht immer vor und ein Schluss von den faktisch beobachteten Abläufen zu allen möglichen Abläufen ist prinzipiell problematisch. Vor diesem Hintergrund scheint ein Maß der zweiten Gruppe zur Bestimmung der Ähnlichkeit der methodischen Ordnung eine sinnvolle Wahl zu sein. Sinnvolle Ergebnisse konnten mit dem Maß *causal footprints* [7] erzielt werden, da dieses die methodische Ordnung der Prozessmodelle berücksichtigt und gleichzeitig in annehmbarer Zeit berechnet werden kann.

Ergebnis ist eine Ähnlichkeitsmatrix, welche die Ähnlichkeit zwischen den Ausgangsmodellen jeweils paarweise beschreibt.

4.5 Schritt 3: Cluster-Analyse

Das *Ziel* des dritten Schrittes besteht in der Durchführung der Cluster-Analyse. Grundsätzlich können hierarchische und nicht-hierarchische Cluster-Methoden unterschieden werden [8]. Die Verwendung hierarchisch-agglomerativer Verfahren hat den Vorteil, dass die hierarchische Cluster-Struktur unmittelbar als Anhaltspunkt zur schrittweisen Integration der Ausgangsmodelle verwendet werden kann. Derartige Informationen haben sich bei der Beurteilung eines sinnvollen Integrationsgrads der Ausgangsmodelle als nützlich erwiesen und liegen bei anderen Cluster-Verfahren nicht vor.

Ein wichtiger Parameter hierarchisch-agglomerativer Cluster-Verfahren ist die Wahl eines geeigneten Fusionsverfahrens für Cluster. Vor dem Hintergrund, dass die zu gruppierenden Modelle gerade zu integrieren sind, erscheint es sinnvoll, ein Fusionsverfahren zu wählen, bei dem die Ähnlichkeit zu *sämtlichen* Modellen in einem Cluster berücksichtigt wird (*complete-linkage*). So verringert sich die Gefahr einer unerwünschten Kettenbildung und die Festlegung einer geeigneten Cluster-Anzahl wird erleichtert.

Das *Ergebnis* dieses Schrittes ist eine Cluster-Struktur.

4.6 Schritt 4: Modellintegration

Das *Ziel* des Schrittes ist die Integration der Ausgangsmodelle. Es wird in mehreren Iterationen vorgegangen. Die im jeweiligen Integrationsschritt zu integrierenden Prozessmodelle ergeben sich direkt aus der zuvor ermittelten Cluster-Struktur. Aufgrund der hierarchischen Cluster-Struktur ist es nicht notwendig, sämtliche in einem Cluster gruppierten Ausgangsmodelle in jedem Schritt zu integrieren. Vielmehr ist es ausreichend, jeweils die bereits teilintegrierten Modelle der beiden zu fusionierenden Cluster zu integrieren.

Die eigentliche Integration zweier Modelle kann dabei manuell durchgeführt werden. Es bietet sich allerdings an, bei diesem Schritt auch auf eine maschinelle Unterstützung zurückzugreifen. Für Details sei auf die Literatur verwiesen [10].

Ergebnis ist eine gemäß der Cluster-Struktur integrierte Menge von Modellen.

4.7 Schritt 5: Interpretation der Cluster und Nachbereitung

Das *Ziel* des letzten Schrittes ist es, einen sinnvollen Integrationsgrad festzulegen. Grundlage für die Entscheidung ist die Komplexität und Verständlichkeit der teilweise integrierten Modelle sowie die jeweiligen Kompressions- und Expansionsraten.

Bei der Wahl des Integrationsgrads sollte zunächst geprüft werden, ob Gruppen identischer oder sehr ähnlicher Modelle identifiziert worden sind. Die Integration derartiger Modellgruppen ist in der Regel unproblematisch. Anschließend ist zu analysieren, inwieweit die Integration weiterer Modelle möglich ist, ohne die Expansionsrate zu sehr ansteigen zu lassen. Die exakte Grenze einer sinnvollen Integration ist abhängig vom Einzelfall. Gleichwohl ist eine Integration regelmäßig fragwürdig, wenn sich die Kompressionsrate wenig bis gar nicht senkt, aber die Expansionsrate im Verhältnis dazu deutlich ansteigt.

Auch wenn die genannten Metriken einen ersten Anhaltspunkt geben, wie die Komplexität und Verständlichkeit der integrierten Modelle zu beurteilen ist, bietet es sich an, die integrierten Modelle auch visuell zu überprüfen. Darüber hinaus sind regelmäßig weitere Faktoren zu berücksichtigen, die nicht zwingend in Prozessmodellen erfasst werden. Zu denken ist beispielsweise an allgemeine organisatorische Rahmenbedingungen oder Kosten und Qualitätsanforderungen der Prozesse.

Im abschließenden Schritt ist zu überlegen, ob weitere Maßnahmen zur Harmonisierung der Ausgangsmodelle sinnvoll sind, da hierdurch die integrierten Modelle vereinfacht werden können [14]. Auch wenn derartige Anpassungen in der Regel nur manuell durchzuführen sind, kann die bestimmte Cluster-Struktur Hinweise für die detailliertere Analyse und mögliche Nachbereitungen liefern.

Ergebnis ist eine sinnvolle und begründete Wahl des geeigneten Integrationsgrades der Ausgangsmodelle und eine entsprechende Menge integrierter Modelle.

5 Implementierung und Machbarkeitsnachweis

Im Rahmen einer prototypischen Implementierung konnte die prinzipielle Machbarkeit des vorgeschlagenen Konzepts nachgewiesen werden (*proof of concept*). Zur Realisierung des Systems wurde auf folgende Komponenten zurückgegriffen:

- *ARIS Plattform 7.1* der Software AG (ARIS): ARIS dient als Werkzeug zur Prozessmodellierung und zur Verwaltung der Prozessbibliothek.
- *ProcessMerger 0.6* von La Rosa et al.: Dieses Werkzeug ermöglicht die Integration von Prozessmodellen [10].
- *System R x64 3.0.3*: Diese Komponente dient zur Durchführung der Cluster-Analyse.
- Werkzeug *RefMod-Miner*: Hierbei handelt es sich um ein eigenentwickeltes Werkzeug für die quantitative Analyse von Unternehmens-, insbesondere Prozessmodellen.

Die zu analysierenden Prozessmodelle können mit ARIS modelliert und gespeichert werden. Zwecks Analyse ist ein Export der relevanten Modelle auf Basis des XML-basierten Austauschformats von ARIS (*ARIS Markup Language*) notwendig. Dieses Format kann vom Werkzeug *RefMod-Miner* eingelesen und verarbeitet werden.

Darüber hinaus stellt das Werkzeug *RefMod-Miner* verschiedene Basisfunktionen zur maschinellen sowie manuellen Erstellung und Verwaltung von Modell-Korrespondenzen und zur Bestimmung von Modell-Ähnlichkeiten zur Verfügung, die für die vorgestellte Methode sinnvoll genutzt werden konnten. Im Rahmen der Implementierung war es notwendig, eine Schnittstelle zum *System R* im Werkzeug *RefMod-Miner* zu realisieren, um entsprechende Verfahren zur Cluster-Analyse nutzen zu können. Im Rahmen der Methode wird die Funktion *hclust* eingesetzt. Auf diese Weise ist es möglich, die Analyse und Integration vieler Prozessmodelle unmittelbar im Werkzeug *RefMod-Miner* vorzunehmen.

Die beschriebenen Funktionen wurden mit verschiedenen Daten getestet. Dabei hat sich gezeigt, dass auf diese Weise nicht nur eine Implementierung der zuvor beschriebenen Methode möglich ist, sondern auch ein produktiver Einsatz realistisch erscheint.

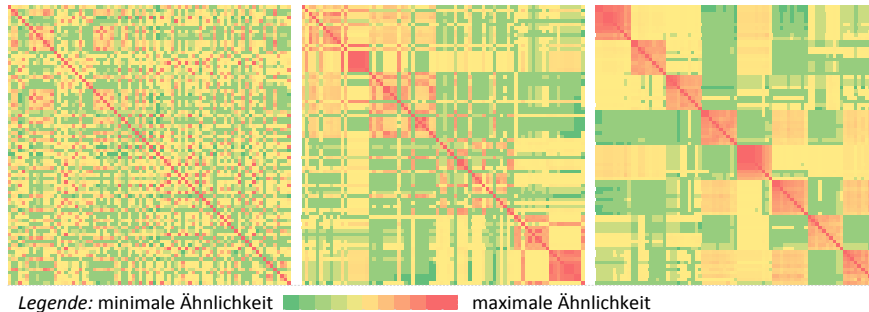


Abbildung 2. Wärmediagramme zur Visualisierung der Modellähnlichkeit, zufällige Sortierung (links), Sortierung nach Komplexität (Mitte) und Ergebnis der Cluster-Analyse (rechts)

6 Experimentelle Anwendungen

Das prototypisch implementierte Werkzeug wurde im Rahmen verschiedener experimenteller Anwendungen getestet. *Szenario 1* basiert auf den bereits in Abschnitt 2 beschriebenen Daten des Fallbeispiels.

In *Schritt 1* sind die Modelle digital zu erfassen und geeignet zu konvertieren, da sie ursprünglich in der Modellierungssprache *YAWL* vorliegen. Die Konvertierung der Modelle erfolgt gemäß der von Vogelaar et al. beschriebenen Konvertierungsregeln [17]. Eine weitere Harmonisierung der Prozessbeschriftungen ist nicht notwendig, da bereits bei der ursprünglichen Erhebung und Erstellung der Prozessmodelle Namenskonventionen vorgegeben und eingehalten worden sind.

In *Schritt 2* wurde die Ähnlichkeit zwischen den Prozessmodellen bestimmt. Hierfür wurde das Maß *causal footprints* verwendet [7], das insbesondere auch die methodische Ordnung der in einem Prozessmodell repräsentierten Aktivitäten berücksichtigt. Abbildung 2 zeigt drei Wärmediagramme (englisch *heatmap*), die Ähnlichkeiten zwischen den Modellen der Modellbibliothek visualisieren. Ein solches Wärmediagramm besteht aus 80 mal 80 Punkten, die jeweils die Ähnlichkeit eines Modellpaares der Bibliothek repräsentieren. Aufgrund der Symmetrieeigenschaft der Ähnlichkeitsbeziehung sind die Diagramme ebenso symmetrisch zur Hauptdiagonalen. Die drei Wärmediagramme beruhen auf einer unterschiedlichen Sortierung der Ausgangsmodelle. Während das linke Diagramm eine zufällige Sortierung der als Punkte repräsentierten Modelle verwendet, sind die im mittleren Diagramm repräsentierten Modelle hinsichtlich der Komplexität gemessen in Anzahl der Knoten und Dichte aufsteigend sortiert. Das rechte Diagramm nutzt eine Sortierung, die aus der im nächsten Schritt durchgeführten Cluster-Analyse hervorgeht.

Abbildung 3 veranschaulicht in Form eines Dendrogramms die in *Schritt 3* im Rahmen der Cluster-Analyse erhaltene Cluster-Struktur. Das entsprechende Wärmediagramm zeigt graphisch, dass innerhalb der Modellbibliothek eine deutliche Struktur

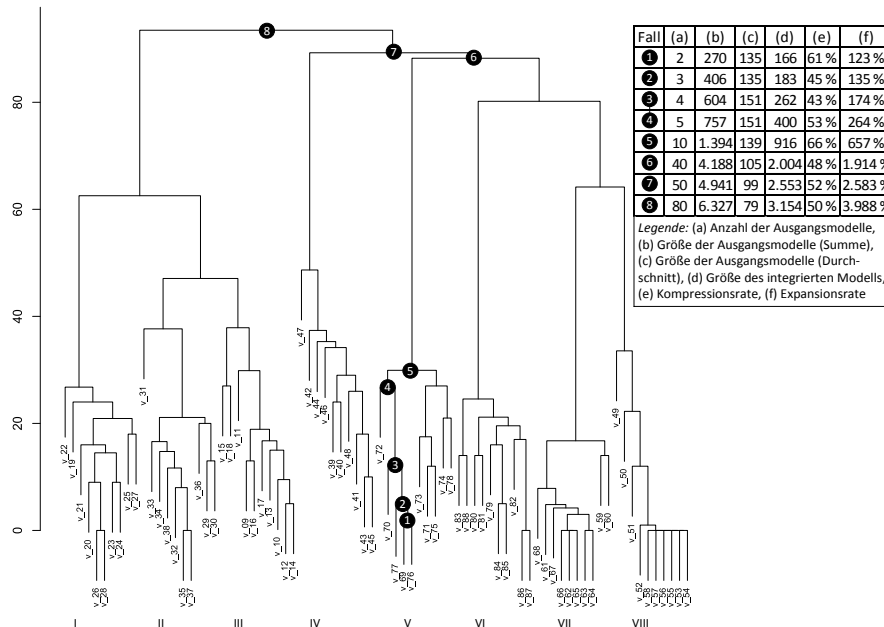


Abbildung 3. Dendrogramm der Cluster-Analyse

vorhanden ist, die im Rahmen der Modellintegration geeignet genutzt werden kann. Im *Schritt 4* wurde eine entsprechende Integration der Modelle vorgenommen. Die Tabelle in Abbildung 3 führt exemplarisch für einzelne Integrationschritte entsprechende Kennzahlen an, aus denen deutlich die vermuteten Effekte zu erkennen sind. Interessant ist, dass die Kompressionsrate nicht monoton fällt, sondern teilweise wieder ansteigt.

In *Schritt 5* ist ein geeigneter Integrationsgrad zu wählen. Aus der von Vogelaar et al. gegebenen Beschreibung [17] lässt sich schließen, dass die analysierte Prozessbibliothek Modelle von acht verschiedenen Prozessen in jeweils zehn Varianten enthält. In Abbildung 2 (rechts) ist diese Domänenstruktur klar visuell ersichtlich und kann auch anhand des Dendrogramms in Abbildung 3 eindeutig rekonstruiert werden (Gruppen I bis VIII). Gleichwohl muss eine mögliche Integration differenzierter betrachtet werden. So zeigt beispielsweise die Analyse der in Abbildung 3 näher beschriebenen Integrationsfälle 1 bis 8, dass sich die Größe des integrierten Modells bereits nach der Integration der fünften Prozessvariante weit mehr als verdoppelt hat. Inwieweit dennoch eine weitere Integration sinnvoll ist, wäre zu hinterfragen. Anders dagegen ist die Struktur bei der Gruppe VIII: Hier ist eine homogene Untergruppe mit sechs Prozessvarianten deutlich ersichtlich, die ohne einen größeren Expansionseffekt integrierbar ist. Erst die Integration der vier weiteren Prozessvarianten lässt die Größe des integrierten Modells deutlich ansteigen. Inwieweit die Qualität der integrierten Modelle durch

manuelle Nacharbeiten weiter gesteigert werden kann, bleibt zukünftigen Analysen überlassen.

Grundlage für das *Szenario 2* sind die Modelle des Wettbewerbs zur automatisierten Identifikation von Korrespondenzen in Prozessmodellen, der im Rahmen der *International Conference for Business Process Management 2013* durchgeführt wurde (*The Process Model Matching Contest 2013* [5]). Die 18 im Wettbewerb genutzten Prozessmodelle repräsentieren zwei unterschiedliche Domänen: Immatrikulation an einer Universität (*university admission*, durchschnittlich 22 Funktionen) und Anzeige einer Geburt in verschiedenen Ländern weltweit (*birth registration*, 18 Funktionen).

Die Prozessmodelle in diesem Szenario sind im Unterschied zu den Modellen aus dem ersten Szenario *nicht* harmonisiert, das bedeutet insbesondere: Erstens kann aus der Gleichheit der Bezeichner nicht geschlossen werden, dass die Knoten dieselbe Funktion beziehungsweise dasselbe Ereignis repräsentieren. Zweitens kann aus der Verschiedenheit der Bezeichner nicht geschlossen werden, dass die Knoten verschiedene Funktionen beziehungsweise Ereignisse repräsentieren. Aufgrund dieser Besonderheit besitzt das zweite Szenario eine höhere Komplexität als das erste Szenario, auch wenn es im Hinblick auf die Anzahl der zu integrierenden Modelle eine geringere Komplexität aufweist.

Zunächst sind die Modelle in *Schritt 1* zu harmonisieren. Hierfür war es notwendig, die in der *Petri Net Markup Language* (PNML) repräsentierten Modelle mithilfe von *ProM* in die *EPC Markup Language* (EPML) zu transformieren; dieses Format kann unmittelbar vom Werkzeug *RefMod-Miner* eingeladen werden. Die weitere Identifikation von Modellkorrespondenzen wird mithilfe der im *ProcessMerger* implementierten Techniken vorgenommen, die sowohl syntaktische als auch semantische Aspekte umfassen. Der Schwellenwert für die Bestimmung der Knotenähnlichkeit wurde einheitlich auf 0,75 festgesetzt. Die folgenden *Schritte 2 bis 4* wurden analog zum ersten Szenario ausgeführt. Die in *Schritt 5* vorgenommene Analyse der integrierten Modelle zeigt, dass die beiden Domänen deutlich getrennt und die entsprechenden Cluster erst auf oberster Stufe fusioniert werden. Gleichwohl lässt sich eine Struktur innerhalb der Modellvarianten erkennen, die bei manueller Überprüfung bestätigt werden kann. Beispielsweise weisen die Immatrikulationsprozesse in Potsdam, Erlangen und Berlin sowie die Prozesse zur Geburtsanzeige *p246* bis *p249* eine sehr hohe Ähnlichkeit auf und können jeweils verhältnismäßig unproblematisch integriert werden.

Weiterhin können die vermuteten Expansions- und Kompressionseffekte nachgewiesen werden. Im Verhältnis zum ersten Szenario fällt der Expansionseffekt ausgeprägter aus, sodass ein höherer Integrationsgrad ohne eine manuelle Nachbereitung weniger sinnvoll erscheint.

7 Diskussion und Ausblick

Die Machbarkeit des vorgeschlagenen Konzepts zur Integration vieler Prozessmodelle wurde nachgewiesen. Zudem konnten im Rahmen der experimentellen Anwendung nützliche Erkenntnisse für die Integration zweier Prozessbibliotheken gewonnen wer-

den. Dabei kann je nach Anwendungskontext ein höherer oder niedriger Integrationsgrad sinnvoll sein, der zweckgerichtet und flexibel wählbar ist. Folgende Punkte sind zu diskutieren:

Wahl des Cluster-Verfahrens: In dem Beitrag wurde ein hierarchisch-agglomeratives Verfahren gewählt, das eine hierarchische Cluster-Struktur liefert und damit im Hinblick auf die Integration der Ausgangsmodelle nützlich ist. Dieses Verfahren konnte zudem die in beiden Szenarien erwartete Gruppen-Struktur reproduzieren.

Wahl des Fusionsverfahrens: Voruntersuchungen zu der Wirkung alternativer Fusionsverfahren haben gezeigt, dass Cluster auf unterschiedlichen Ähnlichkeitsniveaus fusioniert werden. Indes haben sich die Cluster-Strukturen als Ganzes sehr robust gezeigt.

Wahl des Ähnlichkeitsmaßes: Die experimentellen Anwendungen haben gezeigt, dass Maße, die ausschließlich die Ähnlichkeit der gemeinsamen Prozessschritte berücksichtigen, im Hinblick auf die Integration zu Verschlechterungen führen. Gleichwohl ist auch die Nutzung alternativer Ähnlichkeitsmaße intensiver zu erforschen.

Überprüfen und Aufdecken der Domänenstruktur: Mithilfe der Methode wird es möglich, vorhandenes Vorwissen zu überprüfen und weiter zu verfeinern, indem beispielsweise innerhalb einer a priori bekannten Gruppe von Prozessvarianten Untergruppen oder „Exoten“ identifiziert werden.

Alternative Integrationswege: Auch wenn die vorliegende Untersuchung ihren Schwerpunkt in der Integration im Sinne des *Vereinigens* hat, können die vorgestellten Konzepte auch im Hinblick auf eine Integration im Sinne des *Verbindens* eingesetzt werden [14]. In diesem Zusammenhang wäre es wichtig, Ähnlichkeitsmaße zu nutzen, die nicht die Ähnlichkeit gesamter Prozessmodelle, sondern einzelner Teile erfassen.

Aufdeckung von Abstraktionsebenen: Insbesondere im geschilderten Szenario 2 sind auch unterschiedliche Detaillierungsgrade der Modelle bei der Modellintegration von Bedeutung gewesen. Allerdings betreffen die unterschiedlichen Detaillierungen in diesem Szenario nur ausgewählte Prozessbereiche. Zukünftig ist zu erforschen, inwiefern die vorgeschlagene Methode um bekannte Verfahren zur automatisierten Abstraktion von Prozessmodellen erweitert werden kann.

Werkzeugunterstützung und System-Architektur: Die Anwendung der vorgeschlagenen Methode ist ohne Werkzeugunterstützung praktisch unmöglich. Die gewählte Architektur des entwickelten Werkzeugs hat den Vorteil, dass sie unabhängig von einem bestimmten Modellierungswerkzeug realisiert werden kann. So bestehen nur Grenzen in den gewählten Datenformaten, die im- beziehungsweise exportiert werden können. Auch wenn die gewählte Architektur neue Schnittstellen erfordert, erhöht sie die Flexibilität und Anwendbarkeit des Werkzeugs *RefMod-Miner*.

Zusammenfassend lässt sich festhalten, dass die vorgestellte Methode und die prototypische Implementierung eine leistungsfähige Technik bereitstellen, um die eingangs skizzierten Herausforderungen der Integration von Prozessmodellen im Großen zu unterstützen. Künftige Arbeiten werden auf die weitere Anwendung und Verbesserung der vorgeschlagenen Methode ausgerichtet sein. Dabei wird auch die softwaretechnische Methodenunterstützung weiter ausgebaut, da nur so die Praktikabilität der Methode gewährleistet bleibt. Auf diese Weise werden die Theorie und Technik der Modellintegration innerhalb der Wirtschaftsinformatik erweitert, sodass die modellbasierte Gestaltung von Informationssystemen zunehmend ausreift.

Danksagung: Die vorgestellten Forschungsergebnisse wurden unter anderem im Rahmen des Projektes „Konzeptionelle, methodische und technische Grundlagen zur induktiven Erstellung von Referenzmodellen (Reference Model Mining)“ erarbeitet, das durch die Deutsche Forschungsgemeinschaft gefördert wird (GZ LO 752/5-1).

Literatur

1. Aier, S. et al.: Referenzprozesse empirisch bestimmen – Von Vielfalt zu Standards. *Wirtschaftsinformatik & Management* 3, 14-22 (2011)
2. Alves de Medeiros, A.K. et al.: Quantifying process equivalence based on observed behavior. *Data Knowl. Eng.* 64, 55-74 (2008)
3. Batini, C. et al.: A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys* 18, 323-364 (1986)
4. Becker, M., Laue, R.: A comparative survey of business process similarity measures. *Computers in Industry* 63, 148-167 (2012)
5. Cayoglu, U. et al.: Report: The Process Model Matching Contest 2013. In: Lohmann, N. et al. (Hrsg.) *BPM 2013 Workshops, LNBIP 171*, S. 442-463. Springer, Berlin (2013)
6. Delfmann, P.: Verteilte Informationsmodellierung – Methodische und technische Konzepte zur Disambiguierung und Analyse arbeitsteilig entwickelter Informationsmodelle. *Kumulative Habilitationsschrift*, Münster (2010)
7. Dijkman, R. et al.: Similarity of business process models: Metrics and evaluation. *Information Systems* 36, 498-516 (2011)
8. Everitt, B.S. et al.: *Cluster Analysis*. Wiley, Chichester, UK (2012)
9. Gottschalk, F. et al.: Merging Event-Driven Process Chains. In: Meersman, R., Tari, Z. (Hrsg.) *OTM 2008*. Springer, Monterrey, Mexico (2008)
10. La Rosa, M. et al.: Business Process Model Merging: An Approach to Business Process Consolidation. *ACM Transactions on Software Engineering and Methodology* 22, 11:11-11:42 (2013)
11. Mendling, J., Simon, C.: Business Process Design by View Integration. In: Eder, J., Dustdar, S. (Hrsg.) *BPM 2006 Workshops*, S. 55-64. Springer-Verlag, Berlin, Heidelberg (2006)
12. Nejati, S. et al.: Matching and merging of statecharts specifications. *Proceedings of International Conference on Software Engineering*, S. 54-63 (2007)
13. Pfeiffer, D.: *Semantic Business Process Analysis - Building Block-based Construction of Automatically Analyzable Business Process Models*. Wirtschaftswissenschaftliche Fakultät. Westfälische Wilhelms-Universität Münster, Münster (2008)
14. Rosemann, M.: *Komplexitätsmanagement in Prozeßmodellen - Methodenspezifische Gestaltungsempfehlungen für die Informationsmodellierung*. Gabler, Wiesbaden (1996)
15. Song, M. et al.: Trace Clustering in Process Mining. In: Ardagna, D. et al. (Hrsg.) *Business Process Management Workshops, LNBIP 17*, S. 109-120. Springer, Berlin (2008)
16. Sun, S. et al.: Merging workflows: A new perspective on connecting business processes. *Decision Support Systems* 42, 844-858 (2006)
17. Vogelaar, J.J.C.L. et al.: Comparing Business Processes to Determine the Feasibility of Configurable Models: A Case Study. In: Daniel, F. et al. (Hrsg.) *Business Process Management Workshops, LNBIP 100*, S. 50-61. Springer, Berlin (2012)
18. Widl, M. et al.: Guided Merging of Sequence Diagrams. In: Czarnecki, K., Hedin, G. (Hrsg.) *5th International Conference on Software Language Engineering*, S. 164-183. Springer, Dresden, Deutschland (2013)