

# Sprachbezogener Abgleich der Fachsemantik in heterogenen Geschäftsprozessmodellen

Janina Fengel, Kerstin Reinking

Fachbereich Wirtschaft  
Hochschule Darmstadt  
Haardtring 100  
64295 Darmstadt  
[janina.fengel@h-da.de](mailto:janina.fengel@h-da.de)  
[kerstin.reinking@h-da.de](mailto:kerstin.reinking@h-da.de)

**Abstract:** In Unternehmen bringt die Geschäftsprozessmodellierung über die Zeit Sammlungen unterschiedlicher Modelle hervor. Sind diese zusammenzuführen, erschweren semantische Unterschiede den inhaltsbezogenen Abgleich, obwohl dies Vorbedingung für ihre Integration wie beispielsweise im Falle von Analysen, Unternehmensumstrukturierungen, Fusionen oder Standardeinführungen ist. Neben semantischer Heterogenität bedingt durch die Verwendung verschiedener Modellierungssprachen liegt ein Haupthindernis für automatisiertes Matching von Modellen in der Art der Nutzung der zur Bezeichnung von Modellen und ihren Elementen gewählten natürlichen Sprache und unterschiedlich genutzter Fachsprachen. In diesem Beitrag wird hierzu eine Methode vorgestellt, wie eine Kombination von Ontology-Matching-Verfahren heuristische Unterstützung bieten kann.

## 1 Hintergrund und Motivation

Die Geschäftsprozessmodellierung zur Beschreibung und Gestaltung betrieblichen Geschehens hat in den vergangenen Jahrzehnten stark an Bedeutung gewonnen. In der Unternehmenspraxis entsteht daher häufig der Bedarf existierende Modelle abzugleichen wie in Fällen von Projekten zur Architektur-, Daten- und Prozessintegration, semantischen Konsolidierungsprojekten, Unternehmensfusionen und B2B-Integrationen sowie bei der Einführung von Standards oder Standardsoftware. Zur Zusammenführung von Geschäftsprozessmodellen sind die vorhandenen Modelle bezüglich der Inhaltsbedeutung ihrer Elemente zu vergleichen, um Entsprechungen, Ansatzpunkte, Schnittstellen oder gar Überschneidungen und Redundanzen ermitteln zu können. Das Vergleichen und Verknüpfen heterogener Modelle ist indes eine nicht-triviale Aufgabe, denn selbst Modelle gleichen Typs unterscheiden sich häufig semantisch [BP08]. Allerdings tritt dabei semantische Heterogenität nicht nur im Bereich der Modellierungssprachen auf, sondern typischerweise bei der Auswahl der natürlich- sprachlichen Fachbegriffe, die zur Benennung der Modellelemente verwendet werden [TF07].

Besonders die frei wählbare Fachterminologie behindert eine Integration von Modellen und damit der zugrunde liegenden Daten und Prozesse, umso mehr bei unterschiedlicher Herkunft der Modelle, sei es aus dezentralen Teams, unterschiedlichen Konzernbereichen oder verschiedenen unabhängigen Unternehmen. Die in natürlicher Sprache formulierten Bezeichnungen spiegeln neben der branchenüblichen Fachterminologie auch die jeweilige tradierte unternehmensspezifische Geschäftssprache wider. Existiert kein allgemein gültiges, verbindlich definiertes Vokabular oder Regeln bezüglich deren Anwendung, können sich Modelle darin erheblich unterscheiden. Erschwert werden Abgleiche nicht nur bedingt durch die Problematik verschiedener Inhaltsbedeutungen der verwendeten Bezeichnungen und das Verständnis davon, sondern auch durch unterschiedlich gewählte Begriffe oder Begriffskombinationen zur Bezeichnung von Modellelementen. Liegen gar Namenskonflikte bedingt durch Synonymie oder Homonymie vor, sind Modelle weder manuell noch automatisiert direkt vergleich- und damit integrierbar [BRS96; TF06]. Insbesondere in großen Unternehmen existiert bereits eine Vielzahl an Geschäftsprozessmodellen, die über die Zeit von unterschiedlichen Personen oder dezentral in Gremien mit mehreren Personen, oft sogar anhand unterschiedlicher Vorgaben erstellt wurden, in verschiedenen Modellierungssprachen oder unter Nutzung unterschiedlicher Fachterminologien. Auch wenn der gleiche Sachverhalt modelliert ist, können sich arbeitsteilig erstellte konzeptuelle Modelle erheblich in ihren Bezeichnern unterscheiden, sodass die für ihre Nutzung notwendige Vergleichbarkeit nicht grundsätzlich vorausgesetzt werden kann [BD10]. Dies gilt umso mehr im Falle des Aufeinandertreffens von Modellen aus bisher unabhängig agierenden Unternehmen oder Unternehmensteilen. Daher gilt es vor Aufnahme weiterführender Arbeiten den semantischen Istzustand zu analysieren. Semantische Ambiguität ist aufzulösen, um die Aussagen von Modellen inhaltlich in Bezug bringen und abgleichen zu können, denn erst der Abgleich der Fachsprache erlaubt die Identifikation von sich inhaltlich entsprechenden Modellen und Modellelementen und darauf aufbauend gegebenenfalls weiterführende strukturelle Vergleiche [SM07]. Bisher sind solche Analyseaufgaben überwiegend nur manuell leistbar. Der notwendige Abgleich und die Integration konzeptueller Modelle wie die hier betrachteten Geschäftsprozessmodelle sind heute rein intellektuelle Arbeiten. Liegen gar viele und große Modelle vor, sind diese Aufgaben ohne automatisierte Unterstützung nur mittels großem Ressourceneinsatz zu erfüllen.

Um diese Lücke schließen und das Potential von Rechenleistung zur automatisierten Verarbeitung nutzen zu können, wird nachfolgend eine entsprechende IT-gestützte heuristische Methode vorgestellt. Dieser Ansatz fokussiert auf die Nutzungsphase nach der Erstellung von Modellen, insbesondere auf Fragen der gemeinsamen Verwendbarkeit. Zur Reduktion der Arbeitslast beim bedeutungsbezogenen Abgleich auf Nutzerseite wird dazu die Anwendung von Semantic-Web-Technologien, insbesondere Ontologieverarbeitung, und eine Kombination von Verfahren zur Verarbeitung natürlicher Sprachen auf die Frage der Ermittlung semantischer Ähnlichkeit von Geschäftsprozessmodellen in Kap. 2 beschrieben. Dazu folgt die Vorstellung der Vorgehensweise zur Erschließung und Formalisierung der in Geschäftsprozessmodellen enthaltenen semantischen Information und der dafür benötigten Ontologien sowie in Kap. 3 des entsprechend implementierten Prototypen. Darauf aufbauend wird in Kap. 4 die Anwendung der Methode gezeigt.

Der Beitrag endet in Kap. 5 mit der Vorstellung und der Verbindung zu verwandten Arbeiten sowie in Kap. 6 einer kurzen Schlussbetrachtung und einem Ausblick auf zukünftige Arbeiten.

## 2 Semantische Analyse

Modelle repräsentieren in der Regel abgestimmtes Fachwissen. Dies ist zum einen Wissen über die Beschreibung von Sachverhalten in Repräsentations- bzw. Modellierungssprachen, zum anderen das Fachwissen zu den modellierten Sachverhalten, beschrieben durch die organisationale bzw. Geschäftssemantik. Die Erschließung und Repräsentation dieses Wissens kann durch semantische Analyse vorgenommen werden [Li00]. Auf diese Weise lassen sich die Beziehungen zwischen den Objekte beider Domänen erfassen und darstellen. Prinzipiell kann die Repräsentation und automatisierte Verarbeitung von Wissen zum weiteren Ausbau der Informationsverarbeitung beitragen. Im Geschäftsalltag hat die Allgegenwärtigkeit des Internet als globale Infrastruktur zur hohen Akzeptanz webbasierter Unterstützung elektronischer Geschäftsabwicklung beigetragen. Die Entwicklung der Idee des Semantic-Web und seiner spezifischen Technologien bietet nun weiterführend die Möglichkeit der Nutzung webbasierter Ontologien in ihrer Eigenschaft als explizite Spezifikationen als Mittel zur Wissensstrukturierung und Herstellung semantischer Interoperabilität basierend auf offenen Standards. Das Prinzip der Annotation von Information mit Metadaten erlaubt die Repräsentation von Wissen in strukturierter, maschinenzugänglicher Form aufbauend auf Internettechnologien, lesbar sowohl für Maschinen als auch von Menschen [SBH06]. Insbesondere bietet sich die Nutzung solcher semantischer Technologien in den Fällen an, in denen intellektuelle Arbeitsleistung zu kostspielig ist und wiederkehrend Abgleiche insbesondere für große und heterogene Mengen von Daten und Informationen zu leisten sind [Fr10]. Ziel des fachsprachlichen Abgleich von Geschäftsprozessmodellen ist die Unterstützung der Vorarbeiten zu strukturellen Vergleichen von Modellen, die wiederum von der verwendeten Modellierungssprache beeinflusst werden.

### 2.1 Ontologierstellung und Ontology-Matching

Kernelement des Semantic-Web sind Ontologien. Dies sind im informatiktechnischen Sinne Artefakte und können als konzeptuelle Schemata verstanden werden [AF05]. Im Prinzip sind Ontologien Sammlungen von Definitionen von Elementen und ihren Beziehungen und enthalten ein abgestimmtes Vokabular [DOS03]. Sie formalisieren die Bedeutung von Begriffen. Obwohl bei der Entwicklung von Ontologien dasselbe Problem auftritt wie bei der Erstellung von Geschäftsprozessmodellen, nämlich die Entstehung semantischer Heterogenität durch die Wahl der Modellierungssprachen und der Fachsprache für die Bezeichner für Klassen bzw. Konzepte und Relationen, sind diese bei Ontologien wiederum weiterführend automatisiert nutzbar für Abgleiche. Die Forschung im Themenfeld des Ontology-Matching widmet sich Fragen der Abgleichbarkeit und Auflösung semantischer Ambiguitäten [ES07].

Ontology-Matching-Verfahren unterstützen bei der Klärung der Bedeutung verwendeter Begriffe und dienen damit der Ermittlung der Bedeutung von Aussagen über Sachverhalte bzw. deren Beschreibungen. Ziel ist das Auffinden semantischer Relationen, die sich als Ontology-Mappings ausdrücken lassen. Angewendet auf die Frage der Bestimmung der Ähnlichkeit der Inhaltsbedeutungen von Modellen und ihren Elementen können sie als semantische Korrespondenzen dienen. Dies ermöglicht Aussagen der Art „A aus Ontologie X entspricht B aus Ontologie Y“, die sich als Funktionen beschreiben lassen

$$\text{SemCorr}(e_1) = \{\{e_2 \in O_2 | e_2\}, e_1 \in O_1\} \in [0,1]$$

Diese semantischen Korrespondenzen drücken Äquivalenz oder Ähnlichkeit aus. Für das Abgleichen der in Geschäftsprozessmodellen enthaltenen Geschäftssemantik bieten sich elementbasierenden Ontology-Matching-Verfahren an. Ein umfassender Überblick dazu findet sich in [ES07]. Für weiterführende Nutzung können die Korrespondenzen persistiert werden. Dadurch können die verknüpften Ontologien bestehen bleiben, ohne zusammengeführt werden zu müssen. Dies ist besonders im Hinblick darauf nützlich, dass die zugrunde liegenden Modelle nicht ohne weiteres geändert werden können, sondern aktiv genutzt werden. Bewahrte Korrespondenzen bieten stattdessen die Möglichkeit einer virtuellen semantischen Integration.

## 2.2 Erschließung und Formalisierung der Semantik von Modellen

Existierende Geschäftsprozessmodelle sind nicht-ontologische Ressourcen, aus denen durch Reengineering die Bedeutung der Modellaussage extrahiert und semantisch formalisiert werden kann. Eine solche Wiederverwendung von Modellen und ihre Konvertierung in Ontologien erlaubt ihre weiterführende Verwertung, während sie weiterhin unverändert aktiver Nutzung zur Verfügung stehen. Durch automatisierte Dekomposition und Überführung in Ontologien wird Maschinenzugang zum enthaltenen Wissen hergestellt. Ansatzpunkt für die Erschließung des enthaltenen Wissens ist die Überlegung, dass Modelle Fakten aus zwei Wissensbereichen enthalten. Aus dem Sprachraum der Domänensprache sind Konzepte zur Benennung von Modellen und ihren Elementen herangezogen worden, während die Konzepte der Modellierungssprache zur Beschreibung im Sinne einer Typisierung und Anordnung dieser Konzepte genutzt wurden. In Umkehrung dieses Vorgangs lassen sich Modelle zerlegen, um die jeweils verwendeten Konzepte der Sprachräume zu extrahieren und in Form semantischer Modelle zu erfassen, wie in [FR10] beschrieben. Dabei wird die vorhandene Modellinformation ohne manuellen oder zusätzlichen intellektuellen Aufwand an dieser Stelle erschlossen. Die Ontologien zur Beschreibung des Metamodells liegen in OWL bereits vor und können zur Nutzung des Vorgehens der Modelldekomposition verwendet werden. Somit fallen für den eigentlichen Abgleich keine Vorarbeiten an. Bei der Dekomposition werden Modelle mittels XSLT in zwei Ontologien in OWL DL konvertiert. Dies sind die Modellontologie mit den Bezeichnern des Modellnamens und der Modellelemente und die Modelltypontologie mit dem Modell- und den Modellelementtypen.

Zusammen beschreiben sie das Modell mit seinem Namen und Modelltyp sowie die Modellelemente mit ihren Namen und ihrem Modellelementtyp. Bei der Konvertierung werden alle Modellnamen und Modellelementbezeichnungen ohne weitere Verarbeitung „as-is“ transferiert. Auf diese Weise können vollständige Ausdrücke zur Weiterverarbeitung übernommen werden, denn das Fachwissen bei der Modellierung zeigt sich oft erst in der Kombination von Worten zu häufig genutzten Formulierungen. Ebenso bleibt erhalten, dass gegebenenfalls Konventionen die Vergabe von Elementbezeichnern geleitet haben, sowie die verwendete natürliche Sprache und unterschiedlicher Sprachgebrauch genauso wie Besonderheiten der Domäne. Bei Geschäftsprozessmodellen werden zur Bezeichnung von Ereignissen und Aktivitäten zumeist Ausdrücke bzw. Phrasen bestehend aus mehreren Termen verwendet, die selten einen vollständigen Satz bilden. Bei einem semantischen Abgleich ist daher jeder Term einzeln und in seiner Eigenschaft als Teil der vorliegenden Kombination zu betrachten, denn die Phrasen tragen allein in ihrer Gesamtheit die ihnen zugedachte Bedeutung. Augenfälligster Unterschied bei der Analyse der mittels der hier vorgestellten Methode abgleichenden Modellsammlung war die Unterscheidung zwischen Modellen in deutscher und englischer Sprache. Allerdings zeigte sich, dass zumeist keine Umgangssprache zur Anwendung kam und Formulierungen von Emotionen wie Ironie oder Beschönigungen nicht auftraten. Ebenso wurden nur in geringem Umfang beschreibende Adjektive, Adverbien oder modifizierende Ausdrücke gefunden. Dabei wurde auch sichtbar, dass verschiedene Bezeichnungen desselben Begriffs nicht nur durch unterschiedlichen Sprachgebrauch seitens der Modellier, sondern auch begründet durch die Anforderungen und Beschränkungen der jeweiligen Modellierungssprache anzutreffen sind [BD10].

### **2.3 Semantischer Abgleich der natürlichen Sprache der Bezeichner**

Um die entstandenen Modellontologien, die die Domänensemantik der konvertierten Modelle enthalten, automatisiert miteinander in Bezug zu bringen, können Ontology-Matching-Verfahren angewendet werden. Für ansonsten manuell auszuführende Modellabgleiche kann so automatisierte Unterstützung geboten werden und die Modellelemente, die die Domänensemantik widerspiegeln, können unabhängig von der ursprünglich genutzten Modellierungssprache verglichen werden. Dabei zeigte sich, dass die in Prozessmodellen übliche Benennung von Elementen mit mehreren Termen in einer Phrase wie oben beschrieben durch Name-Matching-Verfahren wie beispielweise Zeichenkettenvergleiche bzw. Nutzung von String-Matching-Metriken allein zu minderwertigen Ergebnissen führt. Dies gilt insbesondere bei Vorliegen von Synonymen sowie im Falle unterschiedlicher Positionen gleicher oder ähnlicher Terme innerhalb der zu vergleichenden Phrasen. Stattdessen galt es, verschiedene Anforderungen zu erfüllen. Wie beschrieben führt unterschiedlicher Sprachgebrauch von Modellierern zur Verwendung von unterschiedlichen Bezeichnern. Daher ist davon auszugehen, dass sich Synonyme in den zu vergleichenden Modellen befinden, die beim Einsatz allein von String-Metriken als nicht übereinstimmend erkannt werden könnten. Stattdessen ist die Auflösung von Synonymen erforderlich. Ebenso ist anzunehmen, dass Bezeichner in semantisch ähnlichen Modellen in verschiedenen Sprachen vorkommen können.

Daher ist es erforderlich, dass mehrsprachige Modelle verarbeitet werden können und informationslinguistische Verfahren abhängig von der jeweiligen Sprache genutzt werden. Da es sich bei den Bezeichnern in Modellen um Phrasen handelt, die keine grammatikalisch vollständigen Sätzen oder gar Texte darstellen, sind allerdings einige bestehende informationslinguistische Verfahren nicht direkt anwendbar. Beispielsweise können solchartige Phrasen kaum sinnvoll einer Part-of-Speech-Analyse unterzogen werden. Um eine der Art der Bezeichner angemessene Behandlung zu ermöglichen, wurden verschiedene Verfahren kombiniert, die nachfolgend im Einzelnen kurz vorgestellt werden.

## **2.4 Informationslinguistische Verfahren**

In den vergangenen Jahrzehnten sind verschiedene natural language processing bzw. informationslinguistische Verfahren entstanden, die sich mit der Verarbeitung natürlicher Sprache in bzw. für Informationssysteme befassen [HL09]. Sie eignen sich daher für das Ontology-Matching auf Elementebene [ES07].

### **2.4.1 Kompositazerlegung**

Begriffe in natürlichen Sprachen können unterschiedlich komplex sein, entweder bestehend aus einem Einzelbegriff oder in Form einer Begriffskombination. Dabei besteht ein Einzelbegriff meist aus einem Wort, eine Begriffskombination aus mehreren begrifflichen Bestandteilen. Im Englischen sind dies häufig Mehrwortbenennungen, im Deutschen dagegen Komposita, d.h. die Verbindung mindestens zweier selbstständig vorkommender Worte zu einem zusammengesetzten Wort [Be05]. Für Kompositabildung erlaubende Sprachen wie das Deutsche ist es sinnvoll, Kompositazerlegung durchzuführen und die einzelnen Bestandteile des Kompositums für den Abgleich zu benutzen [St07]. Dabei ist es bei der Dekomposition von Wichtigkeit, sinnvolle begriffliche Bestandteile herzustellen, um alle Vorkommen eines Suchwortes zu finden. Zur Vermeidung nicht sinnvoller Zerlegung von Mehrwortbegriffen oder unerwünschter Zerlegung von Eigennamen können geeignete Wörterbücher unterstützen [Be05].

### **2.4.2 Disambiguierung durch Auflösung von Synonymie**

Synonyme sind unterschiedliche Bezeichnungen für denselben Begriff. Erscheinungsformen dabei sind unterschiedliche Flexionsformen, verschiedene Schreibvarianten eines Wortes, Varianten in unterschiedlichen Zeichensystemen, Abkürzungen oder Vollformen sowie alternativ nutzbare Begriffe [We01]. Durch die Auflösung von Synonymen kann gewährleistet werden, dass semantische Übereinstimmungen zwischen Begriffen gefunden werden, selbst wenn diese unterschiedlich benannt worden sind, sodass die Abgleichsergebnisse verbessert werden [Be05]. Die Auflösung bzw. Word Sense Disambiguation kann unter Zuhilfenahme eines Thesaurus als Synonymwörterbuch vorgenommen werden [St07]. Ein Thesaurus verknüpft Terme zu begrifflichen Einheiten mit und ohne präferierte Bezeichnungen und setzt sie in Beziehung zu anderen Begriffen.

In solchen Begriffsordnungen werden zumeist Beziehungen wie Synonymie und Ambiguität, Hyponymie und Hyperonymie, Antonymie sowie Assoziation erfasst [SS08]. Zur Erstellung webbasierter Thesauri bietet das W3C SKOS, das Simple Knowledge Organization System [MB09]. Die Nutzung von SKOS erlaubt die Wiederverwendung frei verfügbarer Ressourcen, wie beispielsweise WordNet [Fe98] oder den Standard-Thesaurus Wirtschaft (STW) [Zb10].

### **2.4.3 Behandlung von Stoppwörtern**

Im Information Retrieval werden Wörter, die bei Indexierungen nicht beachtet werden, Stoppwörter bzw. stop words, genannt. Zumeist übernehmen sie syntaktische Funktionen und haben somit keine Relevanz für Rückschlüsse auf den Inhalt eines Dokuments. Im Deutschen wie im Englischen sind dies Artikel, Konjunktionen, Präpositionen oder Pronomina sowie die Negation [Be05]. Gleichwohl sind sie für das Verständnis unerlässlich [Be05]. Die Menge an Stoppwörtern kann domänenspezifisch variieren, da auch Wörter enthalten sein können, die, trotzdem sie Bedeutungsträger sind, nicht verwendet werden sollen, da sie in den meisten Dokumenten vorkommen und somit nicht zur inhaltlichen Differenzierung nützen. Entsprechend bietet es sich für die Frage der Geschäftssemantik in Prozessmodellen an, diese nicht generell zu eliminieren wie vorgeschlagen in [Ko07], sondern domänenspezifisch. Abhängig von der Art von Suchen erlaubt der Verzicht auf die Eliminierung bessere Ergebnisse bei Suchen mit Wortkombinationen [Be08]. Weiterhin ist im Falle von Geschäftsprozessen bei Entscheidungen häufig die Existenz der Negation bei der Suche nach semantisch ähnlichen Elementen von Bedeutung. Insbesondere bei Vorliegen kurzer Phrasen, bei denen ein in der jeweiligen Sprache übliches Stoppwort einen erheblichen Bedeutungsunterschied ausmacht, kann die Stoppworteliminierung zu falschen Ergebnissen führen, wie beispielsweise bei Negationen [St07].

### **2.4.4 Stemming**

Zur morphologischen Analyse bieten sich im Information Retrieval Methoden zur Grundformbildung bzw. Lemmatisierung sowie der Wortstambildung bzw. Stemming an [St07]. Bei der Lemmatisierung wird die grammatische Grund- oder Stammform durch die Rückführung der konkreten Wortform auf einen Wörterbucheintrag ermittelt. Beim Stemming werden morphologische Varianten eines Wortes auf ihren gemeinsamen Wortstamm durch die Entfernung von Flexionsendungen und Derivationsuffixen auf einen gemeinsamen Stamm zurückgeführt, wobei dieser nicht zwingend ein lexikalischer Begriff sein muss. Im Falle des Abgleichs von Prozessmodellen können auf diese Weise Bedeutungsähnlichkeiten zwischen Aktivitäten, egal ob mittels eines substantivierten Verbs oder einer Kombination aus Verb und Substantiv benannt, und Objekten genauer ermittelt werden, da hier nur die Stammformen miteinander verglichen werden. Zudem können unerwünschte Matchings von Suffixen ausgeschlossen werden, da diese vor dem Matching entfernt werden.

## 2.4.5 Vergleich von Zeichenketten

Eine Folge von Zeichen eines definierten Zeichensatzes wird als Zeichenkette bzw. String bezeichnet. Strings sind Zeichensequenzen beliebiger Länge aus einem definierten Vorrat [ES07]. String-Matching-Algorithmen suchen Übereinstimmungen von Zeichenketten. Diese Aufgabe fällt in den verschiedensten Domänen an und hat im Laufe der Zeit zu unterschiedlichen Ansätzen geführt [CRF03]. String-Metriken erlauben die Messung von Ähnlichkeiten zwischen Zeichenketten [SSK05]. Die Levenshtein-Distanz zweier Strings ist die minimal erforderliche Anzahl von Einfügungen oder Entfernungen zur Umwandlung der ersten in die zweite Zeichenkette [Le66]. Die Jaccard-Metrik vergleicht die Ähnlichkeit von Worten innerhalb eines Ausdrucks [Ja12]. Die Jaro-Metrik vergleicht Zeichen und ihre Position innerhalb der Zeichenkette, auch wenn sie einige Positionen voneinander entfernt sind [Ja89]. N-Gramme können zur Fragmentierung von Worten bzw. Zeichenketten verwendet werden [St07]. Der darauf basierende Q-Grams-Algorithmus zählt die gemeinsame Menge von Tri-Grammen in den zu vergleichenden Zeichenketten und eignet sich dadurch für so genanntes approximate string matching [ST95]. Da es bei den Ergebnissen aus den verschiedenen Verfahren große Unterschiede geben kann, ist hier die Auswahl einer passenden Metrik in Abhängigkeit von der Sprache und Funktion der Begriffe zu treffen [SSK05]. Obwohl String-Metriken allein nicht alle Bedürfnisse beim Finden von semantischen Ähnlichkeiten von Bezeichnern erfüllen, haben sie sich trotzdem als nützlich in diesem Feld erwiesen [SSK05]. Liegt keine Synonymie von Termen vor, können sie eingesetzt werden, um semantische Ähnlichkeit aufgrund von Übereinstimmungen von Zeichenketten zu bestimmen. Ein vorher durchgeführtes Stemming kann dabei die Präzision der Ergebnisse erhöhen, denn durch die Reduzierung auf den Wortstamm werden dann beispielsweise Übereinstimmungen zwischen Suffixen nicht bewertet.

## 3 Implementierung

Zur Anwendung der beschriebenen Verfahren wurde prototypisch ein System namens LaSMat implementiert, welches für Language-aware Semantic Matching steht.

### 3.1 Technische Realisierung

Die Realisierung der Komponenten erfolgte in Java. Das System kann als Java-API eingebunden oder über eine prototypische Oberfläche angesprochen werden. Abbildung 1 zeigt das Vorgehen zum Abgleich der Modellontologien in Form eines Sequenzdiagramms. Bei einer Anfrage wird im ersten Schritt ein Abgleich beider Phrasen vorgenommen. Dieser Vergleich erfolgt unidirektional. Liegt vollständige Übereinstimmung vor, wird der Wert 1 als Konfidenzwert und damit angenommene Stärke der gefundenen semantischen Korrespondenz zurückgegeben. Ist dies nicht der Fall, werden die Phrasen in Einzelterme zerlegt und diese miteinander verglichen. Hierbei kommen alle oben vorgestellten Verfahren zum Einsatz, wobei die Kompositazerlegung derzeit nur für die deutsche Sprache durchgeführt wird.

Bei allen Verfahren hat der Nutzer die Möglichkeit zu parametrisieren, indem Gewichtungen für die Ergebnisse der verschiedenen Verfahren gesetzt werden können. Die Gewichtung für Übereinstimmungen von als Stoppwort identifizierten Termen ist konfigurierbar. Zur Auflösung von Synonymen können zur Laufzeit Thesauri im SKOS-Format importiert werden. Standardmäßig eingebunden sind WordNet [Fe98] als lexikalische Ressource generell für die englische Sprache im SKOS-Format [W310] und als wirtschaftsspezifische Ressource der STW, der Begriffe in deutsch und englisch enthält [Zb10]. Daneben ist für die generelle deutsche Sprache eine von uns erstellte SKOS-Version des OpenThesaurus in Benutzung [Na05].



Abbildung 1. Sequenzdiagramm des Language-aware Semantic Matchers

Dabei lässt sich über den Parameter  $s \in [0,1]$  als Synonym-Maß die Gewichtung von Synonym-Matches für die Ergebnisaggregation konfigurieren. Für das Stemming werden die Bibliotheken für die deutsche und die englische Sprache aus dem Snowball-Projekt genutzt [PB11]. Für das String-Matching steht eine Auswahl verschiedener String-Metriken zur Verfügung. Es wird dafür die Java-API SimMetrics genutzt [Ch06]. Für die Gewichtung des Ergebnisses in der Gesamtwertung kann ein entsprechender Wert angegeben werden. Zur Ermittlung des Gesamtwerts der Konfidenzen der gefundenen Korrespondenzen werden aus allen Verfahren die besten Ergebnisse aggregiert. Die Ergebnisse sind Matchinginformationen zu jeder Phrase.

Diese lassen sich im INRIA-Format [Eu06] sowie in einer Alignment-Ontology in einem von uns dafür entwickelten Format abspeichern. Die prototypische Oberfläche ermöglicht eine tabellarische Visualisierung der Ergebnisse, wobei zur Filterung ein Schwellwert für die Stärke der gefundenen Korrespondenzen gesetzt werden kann.

### 3.2 Berechnung der semantischen Ähnlichkeit

Gefundene Korrespondenzen werden als Tupel beschrieben in der Form

$$\langle (e_1, m_1), (e_2, m_2), c \rangle$$

wobei

- $(e_k, m_k)$  der Bezeichner eines Elements einer Modellontologie ist,
- $c$  als Konfidenz die angenommenen Stärke der Beziehung darstellt, ausgedrückt als numerischer Wert zwischen 0 und 1.

Der entwickelte Algorithmus bestimmt einen fuzzy Wert für die Ähnlichkeit zwischen zwei Bezeichnern, wobei 1 Äquivalenz ausdrückt und 0 keinerlei Übereinstimmung bedeutet. Wir definieren die Ähnlichkeit zwischen zwei Bezeichnern als arithmetisches Mittel aller Übereinstimmungen in Relation zur Anzahl der Terme in beiden Bezeichnern mit

$$Sim(e_1, e_2) = \frac{\frac{OverallTermSim(e_1, e_2)}{length(e_1)} + \frac{OverallTermSim(e_1, e_2)}{length(e_2)}}{2}$$

wobei

- $length(e_k)$  die Anzahl an Termen der Bezeichnung  $e_k$  ist, ausgedrückt als  $length(e_k) = Num(t_{e_k})$
- $OverallTermSim(e_1, e_2)$  die Gesamtübereinstimmung zwischen allen Termen zweier Bezeichner.

Für die Berechnung der Gesamtübereinstimmung wird das jeweils höchste Ähnlichkeitsmaß zwischen dem aktuell verglichenem Term und allen Termen des zweiten Bezeichners für die Berechnung herangezogen mit

$$OverallTermSim(e_1, e_2) = \sum_{k=1}^{length(e_1)} \max(Sim(t_{k_{e_1}}, t_{1...n_{e_2}}))$$

wobei

- $Sim(t_k, t_n)$  das Ähnlichkeitsmaß zwischen zwei Termen ist.

Die Bestimmung dieses Ähnlichkeitsmaßes basiert auf der Berücksichtigung verschiedener Werte. Im Falle einer exakten Übereinstimmung ergibt das Ähnlichkeitsmaß

$$Sim(t_k, t_n) = 1$$

Dabei wird allerdings für den Fall, dass es sich bei den übereinstimmenden Termen um Stoppwörter handelt, das konfigurierte Stoppwort-Maß anstelle des Wertes 1 verwendet. Im Fall ( $k \neq n$ ) würde das Ergebnis der Distanzmessung sein, dass keine Übereinstimmung vorliegt oder eine gesonderte Behandlung aufgrund der Distanz zwischen den einzelnen Zeichen nötig wäre [Ja89]. Dabei ist jedoch zu beachten, dass die Distanz zwischen zwei Termen, anders als bei reinen Zeichensequenzen, wie beispielsweise Gencodes, nicht in allen Fällen zur Bedeutungsänderungen führt, sondern trotzdem semantische Ähnlichkeit vorliegt. Dies lässt sich am Beispiel der beiden Bezeichner „check invoice“ und „invoice check“ zeigen, bei denen semantische Ähnlichkeit anzunehmen ist. Allerdings lässt die unterschiedliche Positionierung der Terme innerhalb des Bezeichners das Vorliegen unterschiedlicher Wortarten vermuten. Die Distanz der Terme lässt also auf einen Unterschied schließen, der aber kleiner ist als der bei Distanzen zwischen gleichen Zeichen in einem String [PW97]. Unser Ansatz für ( $k \neq n$ ) wird daher weitergeführt als

$$\frac{Sim(t_k, t_n)}{td}$$

wobei

-  $td$  als „term disorder weight“ eingeführt wird mit einem Wert  $\geq 1$ .

Dies folgt dem Ansatz von McLaughlin zur Behandlung von „disagreeing characters“ bei String-Vergleichen wie angewendet in [PW97], wobei jedoch die tatsächliche Distanz der beiden Terme aus oben genanntem Grund außer Acht gelassen wird. Dieser Wert ist konfigurierbar. Ein hoher Wert verringert daher das Ähnlichkeitsmaß zwischen zwei Termen, die an unterschiedlichen Stellen einer Phrase stehen.

### 3.3 Interpretation der Resultate

Die Ergebnisse des Matchings drücken die Stärke einer ermittelten Korrespondenz als Konfidenzwert zwischen 0 und 1 aus. Bei der Analyse der Ergebnisse durch Domänenexperten zeigte sich allerdings, dass die Ergebnisse in dieser Form nicht intuitiv verständlich sind. Daher wird dazu eine Fuzzyfizierung vorgenommen und beginnend bei 1 für  $c = 1$  die Angabe "exactMatch", für  $1 < c > 0,745$  die Angabe "closeMatch", für  $0,745 < c > 0,495$  die Angabe "relatedMatch" Nutzern präsentiert. Dies unterstützt sie bei der Entscheidung bezüglich weiterführender Arbeiten zu Abgleichen oder Analysen.

## 4 Anwendung

Der Prototyp wurde genutzt, um die Machbarkeit und den Nutzen zeigen zu können für eine Sammlung von insgesamt 1.380 Geschäftsprozessmodellen, die zu gleichen Teilen deutsch- oder englischsprachige Bezeichner ihrer Elemente aufweisen. Es handelt sich dabei um Modelle des SAP-Referenzmodells, verschiedene Modellen aus der Literatur sowie Referenzmodelle entnommen aus E-Business-Standards.

## 4.1 Empirische Evaluation

Es wurden aus dieser Sammlung zufällig acht Modellpaare ausgesucht, zwischen denen Ähnlichkeit vermutet wurde. Dabei waren Modelle unterschiedlichen Typs willkürlich aus EPK, BPMN-Modellen und UML-Aktivitätsmodellen ausgewählt. Dazu wurden die konfigurierbaren Werte wie im Screenshot in Abbildung 2 ersichtlich gesetzt.



Abbildung 2. Screenshot des LaSMat

Zur Beurteilung des Ergebnisses der vorgenommenen Abgleiche der Modellontologien, die die Geschäftssemantik repräsentieren, wurden die gefundenen Korrespondenzen mit einer Stärke größer 0,5 verglichen mit Korrespondenzen, die manuell von Domänenexperten als Referenz erstellt wurden. Augenfällig war dabei der Zeitaufwand. Während die menschliche Arbeit für alle ausgewählten Modellpaare bei einem Umfang von einer bis mehreren Stunden lag, dauerte der Abgleich im LaSMat-System zwischen 290 ms bis maximal 3.100 ms pro Paar. Zur Beurteilung der Ergebnisgüte wurde auf Maße aus dem Information Retrieval zurückgegriffen [St07]. Dies sind Precision ( $P$ ), Recall ( $R$ ) und F-Measure ( $F$ ) ausgedrückt als Wert zwischen 0 und 1.  $P$  beschreibt die Korrektheit als Verhältnis aller korrekt gefundener zur Menge aller gefundenen Korrespondenzen.  $R$  beschreibt die Vollständigkeit als Verhältnis aller korrekt gefundenen zur Menge aller erwarteten Korrespondenzen. Zur Gesamtbeurteilung zeigt  $F$  das gewichtete harmonische Mittel dieser beiden Werte. Die Anwendung der Methode ergab für  $P$  einen Mittelwert von 0,89, für  $R$  einen Mittelwert von 0,9 und für  $F$  einen Mittelwert von 0,89. Aus den Mittelwerten der Stichprobe lässt sich für die Grundgesamtheit als Indiz für die Machbarkeit der Methode vermuten, dass bei 5%-iger Irrtumswahrscheinlichkeit die Precision zwischen 0,8 und 0,98 und der Recall zwischen 0,83 und 0,97 liegt, wobei der Maximalwert jeweils 1 ist.

## 4.2 Detailbetrachtung zur Verfahrenskombination

Zur Betrachtung zur Wirkung der Parametrisierung der verschiedenen genutzten Verfahren wurde eine Detailbetrachtung an Einzelbeispielen vorgenommen. Durch die Kompositazerlegung wurden die Ergebnisse erwartungsgemäß verbessert, beispielsweise wurde die Ähnlichkeit zwischen „Rechnungsprüfung“ und „Rechnung prüfen“ ohne Zerlegung mit einem Wert von 0,54 zurück gegeben und mit Zerlegung von 0,74. Synonym-Matches können unterschiedlich gewichtet werden. Dies erscheint sinnvoll in den Fällen, in denen es aufgrund von Quasi-Synonymen zu Bedeutungsverschiebungen kommt. Während das Matching ohne Synonymauflösung keine Übereinstimmung zwischen bedeutungsgleichen Benennungen findet, werden durch die Synonymauflösungen diese Übereinstimmungen gefunden. Dabei führt ein Wert von 0 zu einem Abgleich ohne Synonymauflösung, während alle Werte größer 0 das Ergebnis gewichten. Ein zwischen Stoppworten gefundener exakter Match beeinflusst maßgeblich das Gesamtergebnis beim Phrasen-Matching aufgrund der im Vergleich zu Volltexten geringen Anzahl an Termen. Unser Ansatz, Stoppwort-Matches mit 0.0 zu gewichten, sodass Stoppwort-Matches nicht in die Gewichtung bei der Gesamtähnlichkeitsbewertung fallen, liefert ähnliche Ergebnisse wie die Stoppworteeliminierung, berücksichtigt aber weiterhin die Fälle, in denen ein Stoppwort einen Bedeutungsunterschied ausmacht. Durch Stemming konnten Abgleiche unterstützt werden, wobei für die flexionsstarke deutschen Sprache die Ergebnisse nur in geringerem Umfangs verbessert wurden im Vergleich zum Englischen. Für den Zeichenkettenvergleich kam bei der Evaluation Q-Grams zum Einsatz mit einem Term Disorder Weight von 3 gemäß des Ansatzes von McLaughlin wie oben beschrieben. Dies lieferte unter Beachtung der Position eines Terms innerhalb der Phrase erhöhte Trefferquoten.

## 5 Verwandte Arbeiten

Aufgrund der großen Bedeutung der Modellierung zur Beschreibung und Gestaltung betrieblichen Geschehens kommen in der Folge dem Modellabgleich und der Modellintegration eine immer entscheidendere Bedeutung für die Prozess- und IT-Optimierung und damit letztendlich für die Wettbewerbsfähigkeit von Unternehmen zu. Allerdings liegen trotz dieser Bedeutung keine für den Unternehmenseinsatz geeigneten Methoden und Werkzeuge vor. Einige in der Literatur vorliegende Arbeiten zur Modellintegration konzentrieren sich auf den Bereich der Modellierungssprachen und die Möglichkeiten der Migration oder Integration basierend auf der Übertragung der Modelle von einer Modellierungssprache in eine andere [Ge07; MK07]. Dabei wird der Aspekt heterogen verwendeter Fachsprache nicht betrachtet, sondern die Modellelementbezeichnungen werden unverändert weiter genutzt. Obwohl die Nutzung von Ontologien langfristig als Möglichkeit zur Herstellung eines einheitlichen, gemeinsamen, ständig aktuellen und kollaborativ weiterentwickelten digitalen Modells des ganzen Unternehmens gesehen werden [Fr10], existieren bisher keine Vorschläge zu ihrer Anwendung für Modellabgleiche nach deren Erstellung bzw. für Integrationen oder Konsolidierungen. Existierende Vorschläge zur Integration von Prozessmodellen konzentrieren sich zumeist auf die Phase der Ersterstellung von Modellen.

Dabei wird das Vorliegen eines separat erstellten Domänenmodells zur Bezeichnung von Modellelementen oder für ihren Abgleich vorausgesetzt [BEK06; We07]. Im Gegensatz dazu erfordert unsere Methode keine zusätzlichen Vorarbeiten dieser Art. Andere Ansätze erfordern manuelle Annotationsarbeiten zur Auszeichnung von Prozessmodellelementen zur Ermöglichung semantischer Verarbeitung [HLD07; TF09; BD10]. Aktuell liegen keine Ansätze vor, die semantische Abgleiche und existierender Modelle unter Berücksichtigung sowohl der Modellierungs- als auch der genutzten Fachterminologie und verschiedener natürlicher Sprachen bieten. Hier kann unser Ansatz ergänzend wirkend.

## 6 Schlussbetrachtung

Im vorliegenden Beitrag wurde eine Methode zum semantischen Abgleich bereits existierender Geschäftsprozessmodelle mit Hilfe von Semantic-Web-Technologien, insbesondere Ontology-Matching-Verfahren, vorgestellt. Dadurch wird die Fachsemantik in Modellen maschinell erschließbar und durch eine entsprechende sprachbezogene Auswahl, Kombination und parametrisierbare Ergebnisaggregation mehrerer sprachverarbeitender Verfahren automatisiert abgleichbar. Die ermittelten Ergebnisse können Ansatzpunkte für weiterführende Strukturvergleiche und darauf basierende Verarbeitungsschritte wie beispielsweise Konsolidierungen oder Modelländerungen bieten. Dazu wurde das hier vorgestellte System prototypisch implementiert und für den Machbarkeitsnachweis der entwickelten Methode genutzt. Dabei konnte gezeigt werden, dass die gewählte Kombination von Einzelverfahren Nutzern automatisierte Unterstützung bieten kann. Da das System die Parametrisierung von Gewichtungen vorsieht, ist hierzu weiterführende Evaluation bezüglich deren Effizienz geplant, um domänenspezifisch geeignete Kombinationen ermitteln zu können. Ebenso liefert Ontology-Matching (bisher) keine perfekten Ergebnisse. Insbesondere ist für die Fälle, in denen Phrasen numerische, kryptische oder mischsprachliche Begriffe enthalten, noch weitere Forschungsarbeit nötig. Langfristig könnte weiterführende Forschung bezüglich des entstandenen Bedarfs an Block Matching für das Erkennen taxonomischer und mereologischer Zusammenhänge nutzenstiftend sein. Insgesamt hoffen wir, mit unserem Vorschlag die Nützlichkeit der Anwendung von Semantic-Web-Technologien zur Unterstützung beim Abgleich von Geschäftsprozessmodellen gezeigt zu haben.

## Literaturverzeichnis

- [AF05] Antoniou, G.; Franconi, E.; van Harmelen, F.: Introduction to Semantic Web Ontology Languages. In: Reasoning Web. 1st Int. Summer School 2005, Malta, Springer, Berlin Heidelberg, 2005; S. 1–21.
- [BD10] Becker, J. et al.: Ein automatisiertes Verfahren zur Sicherstellung der konventionsgerechten Bezeichnung von Modellelementen im Rahmen der konzeptionellen Modellierung. In: Modellierung 2010, LNI 161, 2010; S. 49–65.
- [Be05] Bertram, J.: Einführung in die inhaltliche Erschliessung. Ergon., Würzburg, 2005.
- [Be08] Beus, J.: Google changes the treatment of stopwords. <http://www.sistrix.com/news/713-google-veraendert-behandlung-von-stopworten.html>, 30.10.2011.

- [BEK06] Brockmans, S. et al.: Semantic Alignment of Business Processes. In: Proc. of the 8th Intern. Conf. on Enterprise Information Systems (ICEIS 2006). INSTICC, Setúbal, 2006; S. 197–203.
- [BRS96] Becker, J.; Rosemann, M.; Schütte, R.: Prozeßintegration zwischen Industrie- und Handelsunternehmen - eine inhaltlich-funktionale und methodische Analyse. In *Wirtschaftsinformatik* 39, 1996; S. 309–316.
- [BP08] Becker, J.; Pfeiffer, D.: Solving the Conflicts of Distributed Process Modelling – Towards an Integrated Approach. In: 16th Europ. Conf. on Information Systems (ECIS 2008), 2008; S. 1555–1568.
- [Ch06] Chapman, S.: SimMetrics: Open source library of Similarity Metrics. <http://sourceforge.net/projects/simmetrics/>, 17.10.2011.
- [CRF03] Cohen, W.; Ravikumar, P.; Fienberg, S.: A Comparison of String Distance Metrics for Name-Matching Tasks. In: Proc. of IJCAI-03 Workshop on Information Integration on the Web (IIWeb-03), 2003; S. 73–78.
- [DOS03] Daconta, M. C.; Obrst L. J.; Smith K. T.: *The Semantic Web*. Wiley, 2003.
- [ES07] Euzenat, J.; Shvaiko, P.: *Ontology Matching*. Springer, Berlin, 2007.
- [Eu06] Euzenat, J.: An API for ontology alignment. <https://gforge.inria.fr/docman/view.php/117/251/align.pdf>, 17.10.2011.
- [Fe98] Fellbaum, C. Hrsg.: *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, 1998.
- [FR10] Fengel, J.; Rebstock, M.: Domänensemantik-orientierte Integration heterogener konzeptueller Modelle. In: *Modellierung betrieblicher Informationssysteme. Modellgestütztes Management; (MobIS 2010 )*; LNI P-171, 2010; S. 63–78.
- [Fr10] Frank, U.: Interview mit Rudi Studer zum Thema „Semantische Technologien“. In *Wirtschaftsinformatik* 52, 2010; S. 49–52.
- [Ge07] Gehlert, A.: *Migration fachkonzeptueller Modelle*. Logos-Verl., Berlin, 2007.
- [HL09] Harms, I.; Luckhardt, H.-D.: *Virtuelles Handbuch Informationswissenschaft*. <http://is.uni-sb.de/studium/handbuch/>, 30.10.2011.
- [HLD05] Hepp, M. et al.: Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management. In: Proc. of the IEEE Intern. Conf. on e-Business Engineering. ICEBE 2005, IEEE., 2005; S. 535–540.
- [HP11] HP Hewlett Packard: Jena - A Semantic Web Framework for Java. <http://jena.sourceforge.net/>. 30.10.2011
- [Ja12] Jaccard, P.: The Distribution of the Flora in the Alpine Zone. In *The New Phytologist*, 1912, 11; S. 37–50.
- [Ja89] Jaro, M. A.: Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa. *Journal of the American Statistical Association*, 1989; S. 414–420.
- [Ko07] Koschmider, A.: *Ähnlichkeitsbasierte Modellierungsunterstützung für Geschäftsprozesse*. Universitätsverl., Karlsruhe, 2007.
- [Le66] Levenshtein, V.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. In *Cybernetics and Control Theory*, 1966, 10; S. 707–710.

- [Li00] Liu, K.: Semiotics in information systems development. Cambridge Univ. Press, Cambridge, New York, 2000.
- [MB09] Miles, A.; Bechhofer, S.: SKOS Simple Knowledge Organization System Reference. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>, 20.09.2011.
- [MK07] Murzek, M.; Kramler, G.: The Model Morphing Approach – Horizontal Transformations between Business Process Models. In: Proc. of the 6th Intern. Conf. on Perspectives in Business Information Research - BIR'2007, Tampere, Finland, 2007; S. 88–103.
- [Na05] Naber, D.: OpenThesaurus: ein offenes deutsches Wortnetz. <http://www.danielnaber.de/publications/gldv-openthesaurus.pdf>, 12.10.2010.
- [PB11] Porter, M.; Boulton, R.: Snowball. <http://snowball.tartarus.org/index.php>, 31.10.2011.
- [PW97] Porter, E. H.; Winkler, W. E.: Approximate String Comparison and its Effect on an Advanced Record Linkage System. <http://www.census.gov/srd/papers/pdf/rr97-2.pdf>, 10.08.2011.
- [SBH06] Shadbolt, N.; Berners-Lee, T.; Hall, W.: Semantic Web Revisited. In IEEE Intelligent Systems, 2006, 21; S. 96–101.
- [SM07] Simon, C.; Mendling, J.: Integration of Conceptual Process Models by the Example of Event-driven Process Chains. In: 8. Intern. Wirtschaftsinformatik (WI 2007) Univ.-Verl. Karlsruhe, Karlsruhe, 2007; S. 677–694.
- [SS08] Stock, W. G.; Stock, M.: Wissensrepräsentation. Oldenbourg, München, 2008.
- [St07] Stock, W. G.: Information Retrieval. Oldenbourg, München, 2007.
- [ST95] Sutinen, E.; Tarhio, J.: On Using q-Gram Locations in Approximate String Matching. In: Proc. of the 3rd Ann. Europ. Symposium on Algorithms ESA '95. Springer, Berlin, 1995; S. 327–340.
- [StStKo05] Stoilos, G.; Stamou, G.; Kollias, S.: A String Metric for Ontology Alignment. In: ISWC 2005. Springer-Verlag, Berlin Heidelberg, 2005; S. 624–637.
- [TF06] Thomas, O.; Fellmann, M.: Semantische Integration von Ontologien und Ereignis-gesteuerten Prozessketten. In: Proc. EPK 2006 Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten. CEUR-WS.org, Vol. 224, 2006; S. 7–23.
- [TF07] Thomas, O.; Fellmann, M.: Semantic Business Process Management: Ontology-Based Process Modeling Using Event-Driven Process Chains. In IBIS 2, 2007; S. 29–44.
- [TF09] Thomas, O.; Fellmann, M.: Semantische Prozessmodellierung – Konzeption und informationstechnische Unterstützung einer ontologiebasierten Repräsentation von Geschäftsprozessen. In Wirtschaftsinformatik 51, 2009, S. 506–518.
- [W310] Links to SKOS Data. <http://www.w3.org/wiki/SkosDev/DataZone>, 31.10.2011.
- [We01] Weiss, M.: Automatische Indexierung mit besonderer Berücksichtigung deutschsprachiger Texte. <http://www.ai.wu.ac.at/~koch/courses/wuw/archive/inf-sem-ws-00/weiss/index.html>, 30.10.2011.
- [We07] Weske, M.: Business Process Management. Concepts, Languages, Architectures. Springer, Berlin Heidelberg, 2007.
- [Zb10] ZBW Leibniz-Informationszentrum Wirtschaft: STW Standard-Thesaurus Wirtschaft. <http://zbw.eu/stw/versions/latest/download/about.de.html>, 30.10.2011