



Analyse, Monitoring und Vorhersage von zeitbehafteten Geschäftsprozessen

Masterarbeit an der Universität Ulm

Vorgelegt von:

MARTIN SOMMER

martin.sommer@uni-ulm.de

Gutachter:

Prof. Dr. Manfred Reichert

Dr. Ulrich Kreher

2015

Fassung 1. Juni 2015

© 2015 MARTIN SOMMER

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/de/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Satz: PDF- \LaTeX 2 ϵ

Abstract

Heutige Prozess-Management-Systeme stellen immer mehr Daten über Geschäftsprozesse und deren Historie bereit. Diese historischen Daten besitzen in den meisten Fällen zeitbehaftete Attribute wie Zeitstempel oder zeitbehaftete Datentypen. Diese zeitbehafteten Daten können zur Gewinnung von zeitlichen Informationen genutzt und somit zur Generierung von zusätzlichem Wissen verwendet werden. Durch dieses Wissen über einen Geschäftsprozess und dessen Instanzen können Bewertungen und Aussagen über die korrekte Ausführung vergangener und aktueller Instanzen des Prozess abgeleitet werden. Diese Aussagen und Bewertungen von Prozesskomponenten werden durch verschiedene Monitor-Komponenten überwacht, welche Schlüsse über laufende Instanzen ziehen können. Neben einer zeitlichen sequentiellen Abfolge von einzelnen Elementen, können Wiederholungen und Entscheidungen in einer Instanz auftreten. Diese Entscheidungen und Wiederholungen müssen in den Analyse-Komponente berücksichtigt werden. Durch diese automatisierte Analyse-Komponente für zeitbehafteten Daten, werden Prozesse und Aktivitäten klassifiziert und unter Zuhilfenahme verschiedener Kriterien bewertet. Diese Bewertungen können zum einen durch statistische Verfahren und zum anderen anhand von Data-Mining-Algorithmen bestimmt werden. Hierbei wird die Klassifikation von einzelnen Prozesselementen und Instanzen in verschiedene Bereiche eingeteilt. Die Monitor-Komponente bewertet anhand der analysierten Daten die Instanz und zeigt kritische Ausführungspfade an und bewertet die korrekte Ausführbarkeit der Folgeelemente. Für die Analyse- und Monitoring-Komponente werden in dieser Arbeit verschiedene Verfahren vorgestellt und anhand der Anforderungen bewertet. Diese Verfahren können Abhängigkeiten von zeitlichen Attributen verschiedene Fehlerszenarien erkennen. Diese Erkennung von Fehlern ins kritischen Instanzen eines Geschäftsprozesses können in Abhängigkeit des Monitors behoben werden. Zu den theoretischen Verfahren wird eine Machbarkeitsanalyse für die Analyse- und Monitoring-Komponenten entwickelt und implementiert. Diese Implementierung berechnet und beobachtet reale und simulierte historische Daten von zeitbehafteten Geschäftsprozessen, die aus einer realen oder einer simulierten Prozess-Management-Umgebung in Echtzeit erstellt werden.

Danksagung

Ich danke allen, die ich in meinen Jahren an der Universität kennengelernt habe. Jede Bekanntschaft hat mich weitergebracht und habe dadurch viel Erfahrung für das Leben gesammelt. Ein besonderer Dank geht an Andreas Lanz, der meine Bachelorarbeit, mein Seminar, die Projekte und jetzt die Masterarbeit motiviert und anspruchsvoll betreut hat.

Vielen Dank an Alle!

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 1 |
| 1.1 | Aufgaben und Zielsetzung | 3 |
| 1.2 | Aufbau der Arbeit | 4 |
| 2 | Grundlagen | 5 |
| 2.1 | ADEPT-Metamodell | 6 |
| 2.2 | Zeitaspekte in Prozessen | 8 |
| 2.2.1 | Dauer und Zeitabstände von Aktivitäten | 9 |
| 2.2.2 | Wiederholende Aktivitäten - Schleifen | 11 |
| 2.2.3 | Dauer und Abstände in Instanzen | 11 |
| 2.3 | Logging von Ereignissen | 12 |
| 2.3.1 | Komplettes Logging | 13 |
| 2.3.2 | Minimales Logging | 14 |
| 3 | Related Work | 15 |
| 3.1 | Statistische Verfahren | 16 |
| 3.1.1 | Mittel, Modus, Median und Durchschnitte | 16 |
| 3.1.2 | Quantile | 17 |
| 3.1.3 | Histogramme und CUMSUM | 18 |
| 3.1.4 | Normalverteilungen | 20 |
| 3.2 | Data-Mining Verfahren | 21 |
| 3.2.1 | Clusteranalyse | 21 |
| 3.2.2 | Regression | 22 |

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 3.3 | Analysen auf Prozesse und Prozesselementen | 22 |
| 3.3.1 | Verteilungen einzelner Dauern von Aktivitäten | 22 |
| 3.3.2 | Filter und Statistische Verfahren | 23 |
| 3.4 | Monitoring von Prozessen | 24 |
| 3.4.1 | Konversation zwischen Instanzen und Events | 24 |
| 3.4.2 | Definierte Überwachungspunkte in Prozesse | 25 |
| 3.4.3 | Korrelationen von Ereignissen | 26 |
| 3.4.4 | Berechnen der Intervalle zum Schluss | 27 |
| 3.5 | Zeitmanagement in Prozessen | 28 |
| 3.5.1 | Transformation eines zeitbehafteten Prozessgraphen | 28 |
| 3.5.2 | Transformation eines Prozessgraphen in Anlehnung an CSTNU | 29 |
| 3.6 | Fazit | 30 |
| 4 | Analyse und Darstellung von zeitbehafteten Geschäftsprozessen | 31 |
| 4.1 | ETL-Prozess | 32 |
| 4.1.1 | Extraktion | 33 |
| 4.1.2 | Transformation | 34 |
| 4.1.3 | Loading | 35 |
| 4.2 | Zeitbasierte Histogramme | 35 |
| 4.2.1 | Erstellung von zeitbasierten Histogrammen | 35 |
| 4.2.2 | Ausreißer in Histogrammen | 37 |
| 4.2.3 | Angepasste zeitbasierte Histogramme | 39 |
| 4.2.4 | Darstellung der Histogramme in Prozessen | 39 |
| 4.3 | Zeitbasierte Boxplots | 40 |
| 4.3.1 | Erstellung von zeitbasierten Boxplots | 41 |
| 4.3.2 | Darstellung der Boxplots in Prozessen | 44 |
| 4.3.3 | Probleme | 44 |
| 4.4 | Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften | 46 |
| 4.4.1 | Transformation des Prozesses und seiner Instanzen | 47 |
| 4.4.2 | Dauer und Abstände von Aktivitäten | 48 |
| 4.4.3 | Schleifen in Graphen | 50 |

| | | |
|----------|--|-----------|
| 4.4.4 | Zusammenfügen von verschiedenen BranchIDs | 51 |
| 4.5 | Analyse des Ausführungszeitraumes einer Aktivität | 52 |
| 4.6 | Vergleich und Fazit | 54 |
| 5 | Monitoring und Vorhersagbarkeit von zeitbehafteten Geschäftsprozessen | 55 |
| 5.1 | Zusammenspiel der Komponenten | 56 |
| 5.2 | Online und Offline Monitoring | 57 |
| 5.2.1 | Bewertungsgrundlage von Prozesselementen | 58 |
| 5.2.2 | Online-Monitoring | 61 |
| 5.2.3 | Offline-Monitoring | 62 |
| 5.3 | Vorhersage und online Analyse | 64 |
| 5.4 | Rückwirkendes Monitoring | 66 |
| 6 | Prototyp und Implementierung | 67 |
| 6.1 | Aufbau des Prototyps | 67 |
| 6.2 | Simulator | 68 |
| 6.3 | ETL Prozess - MSQL - JAVA | 69 |
| 6.4 | Analyse Komponente - Zweiter Transformationsschritt | 70 |
| 6.5 | Monitoring Komponente | 71 |
| 7 | Zusammenfassung | 73 |
| 7.1 | Fazit | 73 |
| 7.2 | Ausblick | 75 |

1

Einleitung

Durch den Konkurrenzdruck von Ländern wie China oder Indien ist es für Unternehmen immer wichtiger, sich im Wettbewerb gegen diese Länder zu behaupten. Daher gewinnt die Optimierung von Geschäftsprozessen, das Monitoring zur Echtzeit und Darstellung von Prozesseigenschaften für die Bewertung eines Prozesses in den einzelnen Unternehmen immer mehr an Bedeutung [BH06][SS08].

Um diesem Konkurrenzdruck entgegenzuwirken, haben sich in der Vergangenheit *Prozess-Management-System* (kurz: *PMS*) als praktikable Lösung erwiesen [GHS95]. Diese *PMS* erlauben eine Modellierung von Geschäftsprozessen in Verbindung mit Organisationsstrukturen für Benutzer in Unternehmen und ermöglichen eine einfache und automatisierte Ausführung von modellierten Geschäftsprozessen. Diese Gestaltung wird durch ein *Prozess-Metamodell* beschrieben und modelliert eine Abfolge von

1 Einleitung

Arbeitsschritten, auch Aktivitäten genannt. Diese Aktivitäten können in sequentieller, paralleler und wiederholender Abfolge stattfinden. Während der Ausführung lassen sich diese aus Aktivitäten zusammengesetzten Geschäftsprozesse am Monitor überwachen [McC02]. Bei diesen Ausführungen eines Prozesses, auch Instanzen genannt, fallen Daten und Informationen über die Zeitpunkte der Zustandsänderung einer Aktivität, wie beim Starten oder Beenden an. Diese zeitbehafteten Informationen lassen sich für eine Visualisierung der Zeitaspekte und zur Analyse verwenden [RA08]. Durch diese Analyse lässt sich die Korrektheit des Fortschritts einer Instanz und der gestarteten Aktivität definiert. Hierbei müssen kritische oder zu lang dauernde Aktivitäten erkannt werden. Diese Erkennung der Informationen über den Fortschritt kann für das Monitoring von Geschäftsprozessen und der Prozessaktivitäten verwendet werden [GCC⁺04]. In Echtzeit soll dieses Monitoring von Geschäftsprozessen die Informationen einer Instanz bereit stellen und dem Benutzer Erkenntnis über den momentanen Ablauf im Vergleich zur Historie liefern. Diese Informationen bezüglich der Ausführungen der Historie wird dem Benutzer durch statistische und stochastische Methoden zur Verfügung gestellt, welche dem Benutzer erlauben selbstständig kritische Prozesselemente zu erkennen und Verbesserungen zu erarbeiten. Neben der Optimierung der Benutzerseite müssen moderne Monitore die ausgewerteten Informationen an das PMS zurückgeben, welches versucht, die Fehler die in einer Ausführung geschehen sind zu neutralisieren [LM08]. Somit kann der Konkurrenzdruck aus China oder Indien verringert werden, da weniger Ausführungen fehlschlagen und dadurch weniger Kosten entstehen.

Wie beschrieben, ist das Monitoring und die Darstellung von Geschäftsprozessen in Prozess-Management-Systemen eine wichtige Komponente für die Optimierung der Ausführungsreihenfolgen in Verbindung von Zeitaspekten in Geschäftsprozessen [Bul13]. Ziel dieser Arbeit ist es, Komponenten zu entwickeln, um Geschäftsprozesse mit dem Schwerpunkt der Zeitaspekte zu überwachen und Schlüsse über die zeitlichen Abhängigkeiten der Abstände von Aktivitäten und die Dauer einer Aktivität zu beschreiben. Grundlage für die Modellierung der Geschäftsprozesse ist in dieser Arbeit das *ADEPT2*-Metamodell, das an der Universität Ulm entwickelt wurde [Rei00]. Zudem werden zeitbedingte Erweiterungen des *ADEPT2*-Modells für die Darstellung dieser Geschäftsprozesse verwendet [LWR10]. Diese Geschäftsprozesse werden in einem

Prozesssimulator ausgeführt und durch eine Analyse-Komponente für den Monitor und die Darstellung aufbereitet. In der vorliegenden Arbeit werden diese Komponenten und Verfahren für die Analyse, Darstellung, das Monitoring und die Vorhersage beschrieben.

1.1 Aufgaben und Zielsetzung

Das Erreichen der automatisierten Überwachung und der Darstellung von zeitbehafteten Informationen in Geschäftsprozessen bringt einige Problemstellungen mit sich. Zu Beginn existieren zeitbehaftete Daten und Informationen in verschiedenen Log-Dateien und Formaten. Diese Daten müssen in mehreren Schritten gereinigt, exportiert, transformiert und für den Import in das *PMS* und den Monitor angepasst werden. Neben diesen Informationen aus den Daten-Logs stellt ein *PMS* verschiedene Informationen über einen Geschäftsprozess und der Organisationstruktur in einem Unternehmen bereit. Diese müssen während der Transformation angepasst und mit den Daten aus der Vergangenheit kombiniert und verknüpft werden. Das Ziel dieser Arbeit ist, diesen Schritt effizient, agil und variabel zu lösen. Im ersten Schritt muss die Bereinigung der zeitbehafteten Daten und die Informationsgenerierung aus dem *PMS* betrieben werden. Im nächsten wird die Transformation und die Analyse der Daten und Informationen umgesetzt und beschrieben. Diese Analyse muss hierbei die häufigen Prozesseigenschaften wie Sequenzen, Parallelitäten, Entscheidungen und Wiederholungen berücksichtigen und in der Darstellung der zeitbehafteten Informationen beachten. Diese Prozesseigenschaften sind die elementaren Operationen an die Ausführungsreihenfolgen eines Geschäftsprozess. Neben diesen Ausführungsreihenfolgen von Arbeitsschritten existieren Abstände zwischen Aktivitäten und für eine Aktivität die Dauer. Diese Zeitaspekte von Aktivitäten bilden neben den obigen Prozesseigenschaften die Grundlage für die Überwachung und Darstellung im einem *PMS*. Nach der Berechnung und Erkennung der zeitbehafteten Eigenschaften in einem Prozess werden diese für die Darstellung im *PMS* und für die Monitoring-Komponente aufbereitet. Die Darstellung im *PMS* ermöglicht eine übersichtliche Informationsgewinnung von allen Ausführungen auf der Grundlage der Zeitaspekte. Neben dieser Darstellung der in der Vergangenheit ausgeführten zeitbehafteten Geschäftsprozesse ist das weitere Ziel, kritische Momente in einer Ausführung einer Instanz

1 Einleitung

in Echtzeit zu erkennen und den Benutzer und das System über kritische Zustände zu informieren. Diese Information muss für den Benutzer so aufbereitet werden, dass bekannte und gesellschaftliche Eigenschaften wie Farben oder bekannte statistische Verfahren eingesetzt werden. Im Abschluss soll das Thema des rückwirkenden Monitorings für die Fehlerbeseitigung und der Optimierung dargestellt werden. Zuletzt wird die Vorhersage der zeitbehafteten Prozesselementen in Verbindung mit der Analyse Organisationstruktur vorgestellt.

1.2 Aufbau der Arbeit

Die Arbeit ist wie folgt gegliedert: Zu Beginn, werden einleitende Worte über die Intention, Zielsetzung und Aufbau der Arbeit beschrieben. Nach dieser Einteilung werden im folgenden Kapitel 2 die Grundlagen für die vorliegende Arbeit vorgestellt. Diese Grundlagen beziehen sich auf ein Prozess-Metamodell, beispielhaft vorgestellt durch das *ADEPT2*-Metamodell. Fortgesetzt werden die Grundlagen durch verschiedene Zeitaspekte in Prozessen und relevante Ansätze für das Logging von Aktivitäten und weiteren Prozesseigenschaften. In Kapitel 3 werden bekannte und verwandte Arbeiten zum Thema Analyse, Statistik, Vorhersage, Transformation, Zeitaspekte in Prozessen und Monitoring vorgestellt. Nach diesem groben Überblick behandelt das Kapitel 4 verschiedene Verfahren aus der Statistik für die Analyse und der Darstellung zeitbehafteter Geschäftsprozesse. Hierbei wird neben den statistischen Verfahren auch Transformationen und Anpassungen der Prozessgraphen besprochen. Das Kapitel 5 beschreibt aufbauend auf Kapitel 4 das Monitoring und die Vorhersage von zeitbehafteten Geschäftsprozessen. Durch dieses Monitoring werden Prozesse bewertet und für eine Optimierung der Prozessausführungen aufbereitet. Im folgenden Kapitel 6 wird eine Machbarkeitsanalyse und der Prototyp des Monitors, aufbauend auf den Verfahren aus den Kapiteln 4 & 5, beschrieben. Dieser Prototyp kommuniziert mit einem Prozesssimulator, welcher die Informationen aus dem *PMS* des Herstellers *AristaFlow* bezieht. Im letzten Kapitel wird eine Zusammenfassung über die Arbeit und ein Ausblick auf verwandte Themengebiete gegeben. Außerdem wird eine Schlussfolgerung über die vorgestellten Verfahren gezogen.

2

Grundlagen

In diesem Kapitel werden Grundkenntnisse für diese Arbeit erarbeitet und erörtert. Da diese Arbeit auf Geschäftsprozessen basiert, wird im ersten Abschnitt ein *PMS* und das integrierte Prozess-Management-Metamodell erklärt und darauf eingegangen. Dieses PM-Metamodell basiert auf dem *ADEPT2*-Metamodell der Universität Ulm, welches in der Software *AristaFlow* umgesetzt ist. Diese Metamodell wurde mit Zeitaspekte im nächsten Abschnitt 2.2 erweitert, damit zeitbehaftete Geschäftsprozesse modelliert werden können. Diese Prozesse lassen sich ausführen und eine Instanz davon wird erzeugt. Diese Instanzen und ihre Elemente werden während der Ausführung mitgeloggt. Für dieses Logging der Ereignisse aus den Instanzen existieren verschiedene Herangehensweisen mit unterschiedlichem Aufwand. Diese grundlegenden Verfahren werden in Abschnitt 2.3 erläutert.

2.1 ADEPT-Metamodell

Das ADEPT-Modell [Rei00] ist ein blockstrukturiertes Prozess-Metamodell, welches an der Universität Ulm entwickelt wurde. Dieses Metamodell definiert Konstrukte, die bei fast allen Basis-Modellen zu finden sind. Zu diesen Konstrukten gehören beispielsweise Aktivitäten, Kontrollkanten, Sequenzen, Parallelität und Entscheidungen. Im Weiteren werden die einzelnen Konstrukte und Begriffe die in dieser Arbeit von Interesse sind diskutiert.

Prozess: Ein Prozess auch *Workflow* genannt, ist eine formale Sicht auf eine Folge von Schritten um ein Ziel zu erreichen. Dieses Ziel wird meist durch einen Ablauf von Prozessaktivitäten die verknüpft sind erreicht. Eine Darstellung eines Prozesses und dessen Ablauf ist in Abbildung 2.1 zu sehen.

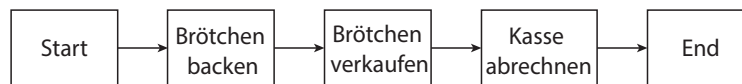


Abbildung 2.1: Einfacher Prozess in ADEPT

Prozessschema: Ein Prozessschema ist die Darstellung eines Prozesses, welches in einem Prozess-Management-System modelliert beziehungsweise ausgeführt wird. Dieses Prozessschema besteht meist aus mehreren Aktivitäten und deren Verbindungen. Zudem werden einzelne Informationen über die Aktivitäten bereitgestellt.

Prozessinstanz: Eine Prozessinstanz, auch *Instanz* genannt, repräsentiert eine Ausführung eines Prozesses. Jede Instanz ist einmalig und hat ihre eigenen Zustände. Eine Prozessinstanz ist vergleichbar mit einer Objektinstanz in der objektorientierten Programmierung.

Aktivität: Eine Aktivität, auch Knoten genannt beschreibt einen Arbeitsschritt. Dieser Arbeitsschritt ist Teil innerhalb eines Prozesses. Aktivitäten können manuell, automatisch oder durch einen Benutzer ausgeführt werden. In ADEPT werden Aktivitäten als Rechtecke dargestellt (siehe Abbildung 2.2 Nummer 1).

Kontrollfluss: Ein Kontrollfluss beschreibt die Abfolge einzelner Aktivitäten in einem Prozess. Diese Abfolge beschreibt bei der Ausführung die Reihenfolge der ausgeführten Aktivitäten. Der Kontrollfluss wird in ADEPT durch Kontrollkanten zwischen Aktivitäten (siehe Abbildung 2.2 Nummer 2), welche als durchgezogener Pfeil dargestellt sind definiert.

Sequenz: Eine Sequenz ist das einfachste Konstrukt in einem Prozess. Sie stellt dabei zwei oder mehrere nacheinander ausgeführte Aktivitäten dar (siehe Abbildung 2.2 Nummer 3). Somit besitzt mindestens eine Aktivität einen Nachfolger.

Parallelität: Eine Parallelität besteht aus zwei oder mehreren Aktivitäten die parallel ablaufen. In ADEPT wird die Parallelität durch einen AND-Block realisiert. Dieser AND-Block besteht aus einem AND-Split und einem dazugehörigen AND-Join (siehe Abbildung 2.2 Nummer 4). Das AND-Split teilt den eingehenden Kontrollfluss in mehrere ausgehende Pfade (auch Zweige genannt) auf. Das AND-Join verbindet alle eingehenden Pfade zu einem ausgehenden Pfad. Die ausgehenden Pfade des AND-Splits und die eingehenden Pfade des dazugehörigen AND-Joins müssen die selbe Anzahl haben.

Entscheidung: Eine Entscheidung, auch XOR-Entscheidung oder Auswahl genannt ist analog wie die Parallelität aus einem XOR-Split und einem dazugehörigen XOR-Join aufgebaut. Dieser XOR-Split und das XOR-Join wird als XOR-Block bezeichnet (siehe Abbildung 2.2 Nummer 5). In einer XOR-Entscheidung wird beim XOR-Split genau ein ausgehender Pfad ausgewählt bzw. alle anderen ausgehenden Pfade werden abgewählt. Wichtig ist es wieder, dass bei dem dazugehörigen XOR-Join die gleiche Anzahl von Pfaden eingehen wie bei dem XOR-Split ausgehen.

Schleife: Eine Schleife ist eine Wiederholung von Sequenzen, Parallelitäten, Entscheidungen und Kombinationen daraus. Ein Schleifen-Block (siehe Abbildung 2.2 Nummer 6) wird durch ein *Start-Loop* gestartet und durch ein *End-Loop* beendet. In ADEPT, wird von dem End-Loopknoten eine Kontrollkante zu dem Start-Loopknoten gezogen und so der Schleifen-Block dargestellt. Bei der Modellierung ist zu Beginn noch nicht bekannt, wie oft die einzelne Schleife wiederholt wird. Bei Erreichen des End-Loopknoten wird jeweils entschieden, ob die Schleife verlassen oder nochmals eine Iteration durchgeführt wird.

2 Grundlagen

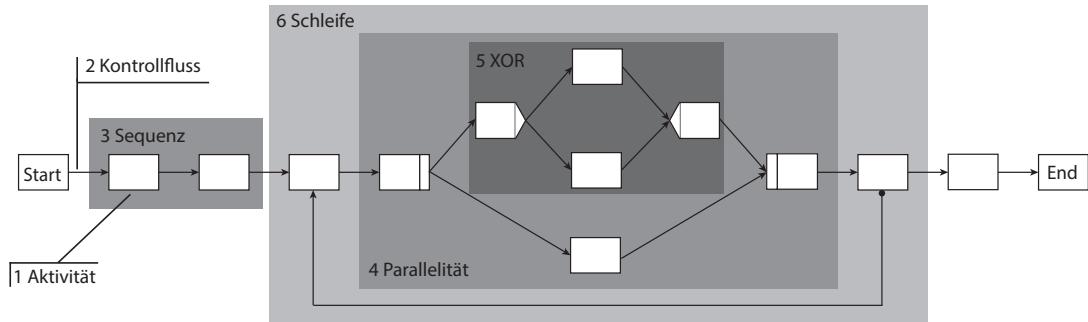


Abbildung 2.2: Basiselemente des ADEPT-Metamodells

2.2 Zeitaspekte in Prozessen

Neben der bisher beschriebenen Kontrollfluss-Sicht, existieren auch weitere Sichten auf Prozesse (z.B. *Daten-Sicht*, *Ressourcen-Sicht* und *Operative-Sicht* [vv02][JB96]). In [LWR10] wurden darüber hinaus die wichtigen Elemente der *Zeitbasierten-Sicht* erfasst und beschrieben. Dabei wurden zehn verschiedene *Zeit-Patterns* (TP) identifiziert, welche in 4 übergeordnete Kategorien einsortiert sind. Diese sind in Tabelle 2.3 dargestellt.

| Pattern Catalogue |
|---|
| Category I: Durations and Time Lags |
| TP1: Time Lags between two Activities |
| TP2: Durations |
| TP3: Time Lags between Arbitrary Events |
| Category II: Restricting Execution Times |
| TP4: Fixed Date Elements |
| TP5: Schedule Restricted Elements |
| TP6: Time-based Restrictions |
| TP7: Validity Period |
| Category III: Variability |
| TP8: Time-dependent Variability |
| Category IV: Recurrent Process Elements |
| TP9: Cyclic Elements |
| TP10: Periodicity |

Abbildung 2.3: Pattern Kategorie [LWR10]

In dieser Arbeit sind hauptsächlich die *Zeit-Patterns* TP1, TP2 und TP9 von Bedeutung. Diese drei Patterns lassen sich einfach aus den Einträgen des Prozess-Event-Logs rekonstruieren und für die Verwendung der Analyse, der Darstellung und für den Monitor nutzen.

2.2.1 Dauer und Zeitabstände von Aktivitäten

In der *Zeitbasierten-Sicht* besitzt eine Aktivität eine minimale und eine maximale Dauer. Diese Dauer wird in Zeiteinheiten wie z.B. Stunden (h), Minuten (min), usw. angegeben. Die maximale Dauer einer Aktivität muss dabei gleich oder größer wie die minimale Dauer sein. Wie in Abbildung 2.4 zu sehen ist, wird eine Aktivität mit einer minimalen Dauer (minD) und maximalen Dauer (maxD) versehen.

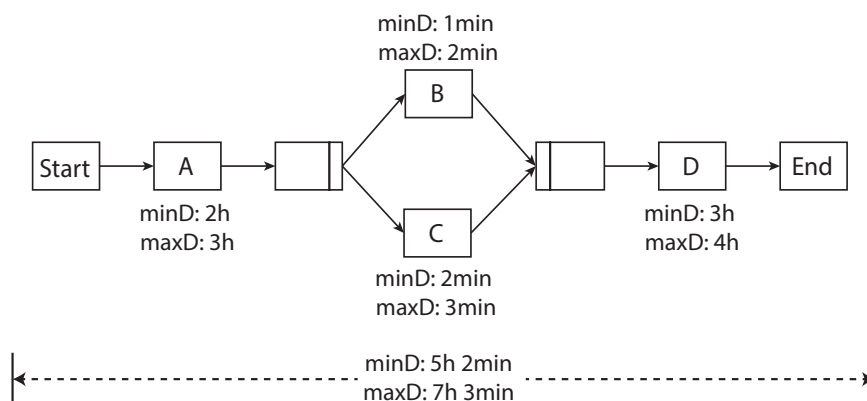


Abbildung 2.4: Aktivitäten- und Prozessdauer

Neben minimaler und maximaler Dauer existieren Zeitkanten, welche zeitliche Abstände zwischen zwei Aktivitäten repräsentieren (siehe Abbildung 2.5). Diese Zeitkanten besitzen jeweils drei Attribute: einen minimalen Abstand, einen maximalen Abstand und die Art der Abhängigkeit der beiden Aktivitäten. Zudem ist die Art der Abhängigkeiten zwischen den Aktivitäten eine der folgenden:

2 Grundlagen

- *Start-Start*: Der zeitliche Abstand ist vom Startzeitpunkt der Startaktivität zum Startzeitpunkt der Zielaktivität definiert. In ADEPT startet die Zeitkante an der linken Kante der Startaktivität und endet an der linken Kante der Zielaktivität.
- *Start-End*: Der zeitliche Abstand ist vom Startzeitpunkt der Startaktivität zum Endzeitpunkt der Zielaktivität definiert. In ADEPT startet die Zeitkante an der linken Kante der Startaktivität und endet an der rechten Kante der Zielaktivität.
- *End-Start*: Der zeitliche Abstand ist vom Endzeitpunkt der Startaktivität und dem Startzeitpunkt der Zielaktivität definiert. In ADEPT startet die Zeitkante an der rechten Kante der Startaktivität und endet an der linken Kante der Zielaktivität.
- *End-End*: Der zeitliche Abstand ist vom Endzeitpunkt der Startaktivität und dem Endzeitpunkt der Zielaktivität definiert. In ADEPT startet die Zeitkante an der rechten Kante der Startaktivität und endet an der rechten Kante der Zielaktivität.

Je nach Abhängigkeit der beiden Aktivitäten ist der minimale bzw. maximale Abstand verschieden definiert. Der minimale Abstand ist die Verzögerung zwischen zwei Aktivitäten und deren Start- bzw. Endpunkten, welcher eintreten muss (z.B. bei *Start-Start* wird der Abstand zwischen den Startzeitpunkten der beiden Aktivitäten definiert). Der maximale Abstand ist die Verzögerung die eintreten kann. Bei Erreichen des Maximums des maximalen Abstandes muss eine Aktivität ausgeführt werden. Analog wie bei der Dauer muss der maximale Abstand größer oder gleich dem minimalen Abstand sein.

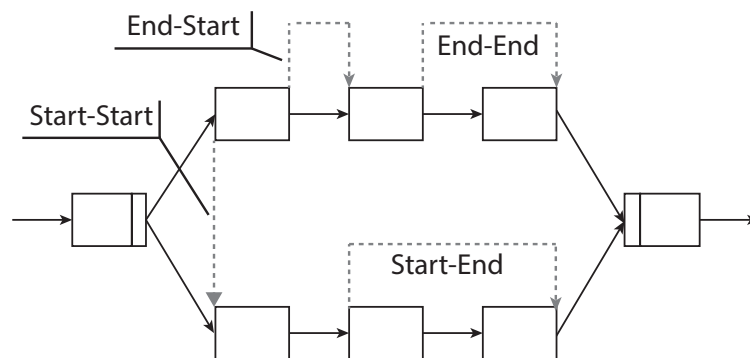


Abbildung 2.5: Prozess mit Zeitkanten in ADEPT

Prozesse besitzen ebenfalls eine minimale und eine maximale Dauer. Diese Dauer setzt sich aus den einzelnen Dauern der Aktivitäten und den zeitlichen Abständen zwischen diesen zusammen. Die maximale Dauer des Prozesses ist der zeitlich schlechteste Weg und die minimale Dauer des Prozesses der zeitlich beste Weg durch einen Prozess. Wie in Abbildung 2.4 zu sehen ist, ist die minimale und maximale Dauer des Prozess von der Startaktivität bis zur Endaktivität gemessen.

Durch die minimale/maximale Dauer einer Aktivität und den minimalen/maximalen Abstand einer Aktivität resultieren für eine Aktivität verschiedene Anfangs- und Endzeitpunkte. Diese allgemeinen Anfangs- und Endzeitpunkte sind in vier Zeitpunkte unterteilt: Den *Frühesten Anfangszeitpunkt (FAZ)*, der den frühesten Moment an dem die Aktivität starten kann beschreibt, den *Spätesten Anfangszeitpunkt (SAZ)*, der den spätesten Moment an dem eine Aktivität starten muss beschreibt, den *Frühesten Endzeitpunkt (FEZ)*, der den frühesten Moment an dem eine Aktivität beendet sein kann beschreibt und den *Spätesten Endzeitpunkt (SEZ)*, der den spätesten Moment an dem die Aktivität beendet ist beschreibt. Diese vier Zeitpunkte sind in der *Zeitbasierten-Sicht* von großer Bedeutung.

2.2.2 Wiederholende Aktivitäten - Schleifen

Bei einer Schleife wird am Ende entschieden, ob die Schleife wiederholt oder ob sie verlassen wird. Findet eine weitere Iteration der Schleife statt, können zeitliche Abstände zwischen Aktivitäten in den verschiedenen Iterationen entstehen. Die Modellierung eines zeitlichen Abstandes zwischen zwei Iterationen einer Aktivität ist in Abbildung 2.6 zu sehen. Wie bei einem zeitlichen Abstand üblich, existiert ein minimaler und maximaler Abstand, sowie die Art der Abhängigkeit (*Start-Start, Start-End, End-Start, End-End*)

2.2.3 Dauer und Abstände in Instanzen

In Abschnitt 2.2.1 wurde auf die Dauer von Aktivitäten und Abständen von Aktivitäten eingegangen. Hierbei wurde dargestellt, dass die Dauern und Abstände ein Minimum

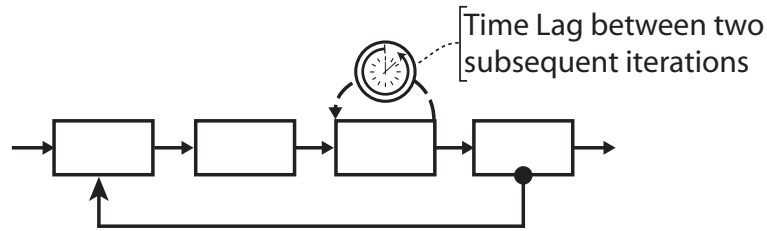


Abbildung 2.6: Weiderholungen bzw. Cycles [LWR10]

und Maximum besitzen. Dieses Minimum und Maximum beschreibt aber nur den theoretischen Zeitaspekt bei der Modellierung. Während der Ausführung eines Prozesses, kann die Dauer oder der Abstand einen beliebigen Wert, auch außerhalb der Modellierung annehmen. Zudem existieren in den wenigsten PMS Einstellungen für die Zeitaspekte. Um die Dauer und einen Abstand zwischen Aktivitäten zu bestimmen, muss man in jeder Instanz die Differenz von dem Ende einer Aktivität und dem Start der Aktivität bilden. Durch diese Differenz erhält man die Dauer der jeweiligen Aktivität in einer Instanz. Durch Bestimmung des Minimum und Maximum aller gleichen Aktivität in einer Instanz des selben Prozesses, lassen sich die Zeitaspekte aus Abschnitt 2.2.1 darstellen. Jedoch können fehlerhafte Messpunkte das reale Ergebnis verschleiern. Analog zu den Dauern können Abstände zwischen zwei Aktivitäten gemessen werden. Diese Abstände können durch die Abhängigkeiten des Abschnitt 2.2.1 beschrieben und bestimmt werden.

Neben den einzelnen Zeitaspekten die in Prozessen dargestellt werden können muss man diese Aspekte erst berechnen und bestimmen können. Dies wird durch das Logging von Ereignissen der Aktivitäten ermöglicht. Im nächsten Abschnitt werden diese Logging Verfahren von Instanzen vorgestellt.

2.3 Logging von Ereignissen

Um Informationen über Prozesse und ihre Zeitaspekte zu generieren, müssen basierend auf den Ereignissen und Zustandsänderungen von Aktivitäten und Instanzen Einträge in den Daten-Log erzeugt werden. Dabei existieren für diese Eintragung in die Log-

Dateien verschiedene Verfahren des Loggings. Hierbei wird unterschieden zwischen dem Aufwand und der Komplexität des Loggings. Manche Logging-Verfahren decken alle Zustandsänderungen von Aktivitäten ab, andere nur das nötigste. Weiter ist die Komplexität der Verzweigungen von Tabellen und die Darstellung von Schleifen ein bedeutender Faktor. Hierbei ergibt sich die Regel, je komplexer die Tabellenstruktur, desto einfacher das Gewinnen der Informationen. Jedoch ergeben sich hierbei eine Verringerung der Geschwindigkeit beim Einfügen und Erzeugen von Einträgen. Beim Logging ist es zudem nötig, dass jeder Eintrag im Daten-Log einmalig und eindeutig identifizierbar ist. Als praktikable Lösung wurden hier UUID (Universally Unique Identifier V4) entwickelt, die durch ihre Erzeugungs-Zeit und einer zufallsgenerierten Zahl des Datentyps *long* erstellt wird. In den folgenden Abschnitten werden zwei Verfahren beschrieben. Eines für das komplette Logging der Zustandsänderungen und eines für die minimale Anforderung, welches die Verfahren aus den Kapiteln 4 und 5 benötigen.

| UUID | Timestamp | ActID | Event |
|--------------------------------------|------------|-------|-----------|
| 5333cf62-e1d3-4981-a8b2-7d4c59cfd1ef | 1433016512 | 1 | STARTED |
| 32a3ed02-f2df-421a-8b55-e5f0ced6e9a3 | 1433059997 | 2 | STARTED |
| 8ef03a7d-fa9b-4ffc-908b-0b54d2838ce6 | 1433016988 | 1 | COMPLETED |
| 7ce617bb-70f2-4ebb-9b2e-62b9766842bb | 1433020030 | 4 | STARTED |

Tabelle 2.1: Prozess-Event-Log

2.3.1 Komplettes Logging

Bei der Ausführung werden die einzelnen Aktivitäten im ADEPT-Metamodell mit einem Zustand bewertet. Dieser Zustand kann sich beispielsweise bei der Aktivierung oder Deaktivierung einer Aktivität ändern. Die einzelnen Zustände und ihre Übergänge sind in Abbildung 2.7 gezeigt. Dabei wurde das Zustandsdiagramm des ADEPT-Metamodells um einen weiteren Zustand *Blocked* erweitert. Dieser Zustand ist für die Zeiterweiterung des ADEPT-Metamodells aus dem vorigen Abschnitt nötig, um das Warten auf die Erfüllung eines minimalen Abstands zu realisieren. Diese Zustände *NotActivated*, *Activated*, *Started*, *Blocked*, *Skipped* und *Completed* werden durch Aktivitäten in einer Instanz angenommen und in diesem Moment im Log gespeichert. Die der Tabelle 2.1 zeigt

2 Grundlagen

einen Daten-Log, mit einer UUID (Version 4) den Erzeugungs-Zeitstempel, die ID einer Aktivität und das dazugehörige Ereignis.

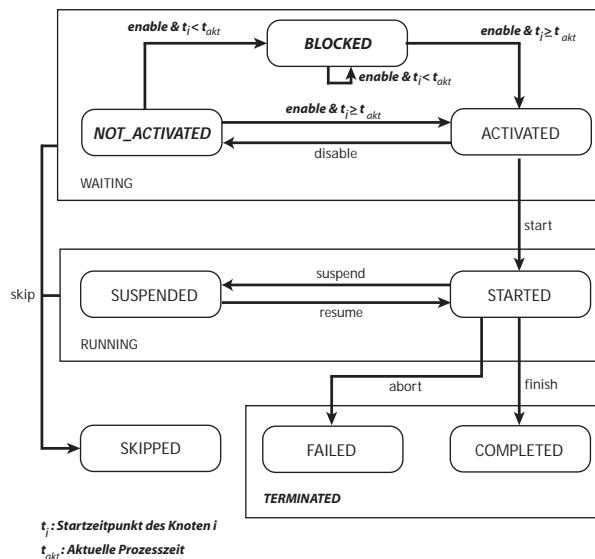


Abbildung 2.7: ADEPT Knotenzustände mit Zeiterweiterung (in Anlehnung an [Rei00])

2.3.2 Minimales Logging

Die minimale Anforderung an das Logging erfüllt die Software AristaFlow. Dieses minimale Logging ist die Grundlage für die Kapitel 4 und 5. Bei diesem Logging werden nur die Zustandsänderungen des *Started* und *Completed* von Aktivitäten beachtet und gespeichert. Durch diese beiden Zustände lassen sich die Dauer einer Aktivität und die Abstände zwischen Aktivitäten bestimmen. Die weiteren Zustände aus Abbildung 2.7 können zum Teils rekonstruieren werden. Diese Rekonstruktion können wir nur durchführen, wenn wir das Logging-Verfahren und die unterstützten Zustände der PMS kennen.

Im nächsten Kapitel werden verwandte Themengebiete aus der Statistik, der Prozessanalyse, der Transformation von Prozessen, weitere Aspekte des Monitorings und Zeitaspekte in Geschäftsprozessen vorgestellt.

3

Related Work

Dieses Kapitel beschreibt verschiedene Herangehensweisen für die einzelnen Komponenten des Monitorings und der Darstellung von zeitbehafteten Geschäftsprozessen. Für die Analyse werden zu Beginn Grundlegende statistische, stochastische und Data-Mining-Verfahren vorgestellt. Diese Verfahren können zeitbehaftete Messreihen von Dauern von Aktivitäten oder Abstände zwischen Aktivitäten beschreiben. Im darauf folgenden Abschnitt 3.3 werden Herangehensweisen für die Analyse von zeitbehafteten Geschäftsprozessen vorgestellt und diskutiert. Zuletzt werden bekannte Verfahren für das Monitoring und Darstellung mit Transformationen von zeitbehafteten Geschäftsprozessen besprochen.

3.1 Statistische Verfahren

Ohne grundlegende statistische Verfahren kann man über Messreihen, die aus der Dauer einer Aktivität oder Abstände zwischen Aktivitäten bestehen nur wenige Informationen extrahieren. Diese statistische Verfahren spielen für die Gewinnung von zusätzlichem Wissen eine große Rolle. Die Statistik ist für den Analyseteil für die Darstellung und Monitoring-Komponente wichtig, da sie automatisiert Daten berechnen und Wissen generieren kann. Neben einfachen Verfahren, wie z.B. die Mittelung oder dem Median von zeitbehafteten Messreihen, werden in diesem Abschnitt Verfahren erläutert, welche die Messreihen grafisch für den Benutzer aufbereiten [Cha83]. Diese Quantile, Histogramme und Verteilungen werden in diesem Abschnitt gezeigt und bezüglich ihrer Festigkeit gegenüber Ausreißern beschrieben.

3.1.1 Mittel, Modus, Median und Durchschnitte

Klassische Berechnungen für die Mittelung von Messreihen sind in der Mathematik Standard-Verfahren in verschiedenen Anwendungsbereichen. Um ein Mittel zu berechnen existieren verschiedene Mittelungsverfahren: das gestutzte Mittel, gleitende Mittel, geometrische Mittel, der Modus (häufigster vorkommender Wert) und der Median (Wert an der mittleren Stelle, einer sortierten Reihe). Die Definition 3.1 zeigt die Definition für das arithmetische Mittel. Weitere Mittel werden durch weitere Verfahren geglättet oder gefiltert. Hierbei sind nur manche Mittel ausreißerneutral und können mit extremen, unerwünschten Werten umgehen. Bei der Berechnung des einfachen Durchschnitts fallen große Ausreißer stark in das Gewicht der Berechnung, wobei bei einem Median extreme Ausreißer, wegen der Teilung der sortierten Messreihe, kaum zu Gewicht fallen. Diese Durchschnitte und Mittel sind die essentielle Grundlage für die meisten Verfahren aus der Statistik und werden in den folgenden Abschnitten benötigt.

$$\bar{x}_{average} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3.1)$$

3.1.2 Quantile

Ein Quantil ist ein Lagemaß, welches eine Messreihe beschreibt. Man teilt hierbei die Messreihe in gleich große Teile. Bei dem kleinsten Quantil wird die Messreihe in zwei Teile geteilt, wobei die Grenze der Median ist. Diese Teilung kann bis in das Endliche weiter betrieben werden. In den meisten Anwendungen werden die Messreihen in Quartile oder in Dezile eingeteilt. Quartile sind hierbei die Grundlage für die grafische Darstellung, der Boxplots.

Quartile (lateinisch „Viertelwerte“) beschreiben die Einteilung von Messreihen in vier gleichgroße Teile. Die Grenzen der Einteilungen werden im 1. Bereich durch den kleinsten Wert der Messreihe und dem unteren Quartil $Q_{0,25}$ (alle Werte die kleiner sind als 25% aller sortierten Werte) bestimmt. Der 2. Bereich ist zwischen dem $Q_{0,25}$ und $Q_{0,5}$ Quartil/Median (Werte zwischen dem $Q_{0,25}$ und kleiner als die Mitte aller sortierten Werte) angesiedelt. Der 3. Bereich liegt zwischen der $Q_{0,5}$ -Grenze und dem oberen $Q_{0,75}$ -Quartil. Zuletzt der 4. Bereich, welcher alle Werte größer der $Q_{0,75}$ beheimatet. Weiter lassen sich, das Minimum/Maximum der Messwerte in dem unteren/oberen Quartil ablesen und die Spannweite der Differenz von Maximum und Minimum bestimmen. Der Interquartilsabstand beschreibt die Ausdehnung, der Differenz des Wertes der $Q_{0,75}$ -Grenze und des Wertes der $Q_{0,25}$ -Grenze.

Dezile (lateinisch „Zehntelwerte“) werden analog wie die Quartile in gleichgroße Teile gesplittet. Dabei werden Messreihen in zehn umfangsgleiche Teile zerlegt. Entsprechend liegt z.B. ein Drittel der Werte unterhalb des dritten Dezils (Quantil $Q_{0,3}$), also 30 % aller Messwerte unterhalb des $Q_{0,3}$. Der Abstand zwischen dem 10%-Dezil und dem 90%-Dezil wird als Interdezilbereich bezeichnet und stellt 80% der Werte dar.

Boxplot

Ein Boxplot basiert auf den Grundlagen der Quartile. Die Messreihe wird in vier umfangsgleiche Teile zerlegt und in einer visuellen Form dargestellt (siehe Abbildung 3.1). Dabei wird zwischen der $Q_{0,25}$ -Grenze und der $Q_{0,75}$ -Grenze eine Box gezogen. Diese Box beinhaltet 50% aller Messwerte und beschreibt durch die Werte der Grenzen den Inter-

3 Related Work

quartilsabstand (kurz IQR). Der Median bzw. die $Q_{0,5}$ -Grenze eines Quartils wird durch einen Strich in der Box markiert und zeigt die Verschiebung der Verteilung (linksschief bei einer Annäherung zur $Q_{0,25}$ -Grenze bzw. rechtsschief bei einer Annäherung an die $Q_{0,75}$ -Grenze) der Messreihe an. Falls eine Normalverteilung vorliegt ist das Boxplot ungefähr symmetrisch.

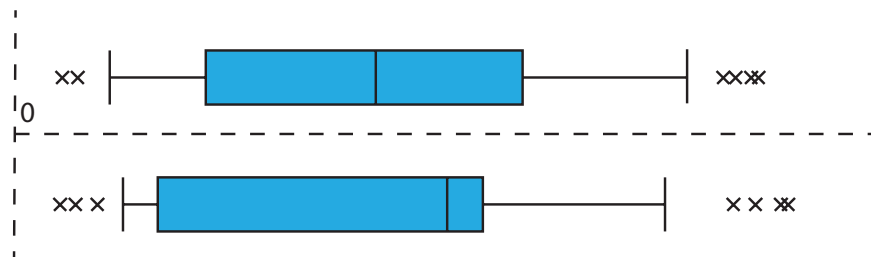


Abbildung 3.1: Boxplot, Antennen und Ausreißer

Der Unterschied zu den Quartilen besteht bei den Boxplots darin, dass Antennen (engl. Whisker) für die Erkennung von Ausreißern hinzugefügt werden. Für die Antennen existieren verschiedene Definitionsgrundlagen. Zum einen die von John W. Tukey stammende Definition [Tuk77], wobei die Länge der untere/obere Antenne den nächsten Wert einnimmt, welcher von der $Q_{0,25}$ - bzw. $Q_{0,75}$ -Grenze 1,5-fachen IQR-Abstand entfernt ist. Innerhalb der unteren und oberen Antennengrenze liegen in den meisten Fällen 96% aller Messpunkte. Die restlichen 4% werden als schwache Ausreißer bezeichnet. Analog zu der Definition von Tukey, existiert eine extremere Darstellung der Antennengrenzen des 3-fachen IQRs. Diese Punkte außerhalb dieser Antennen werden als extreme Ausreißer deklariert und meist in der visuellen Darstellung der Boxplots weggelassen. Diese extremen Ausreißer sind durch die schwachen Antennen bereits abgedeckt.

3.1.3 Histogramme und CUMSUM

Histogramme sind ähnlich wie Boxplots eine visuelle Darstellungsmöglichkeit für die Beschreibung von Messdaten. Jedoch werden die zeitbehafteten Messdaten bei Histogramm-

men meist an einer festen oder variablen skalierbaren Achse in Bereiche eingeordnet und gezählt. Die Breite eines Rechteckes basiert auf den Grenzen des Wertebereichs der Skalierung und die Höhe beschreibt die Anzahl der Häufigkeit aller Messpunkte in diesem Wertebereich (Klassen). Die Anzahl kann sich hierbei auf relative oder absolute Häufigkeitsdichte beziehen. Die Breite der Balken kann durch verschiedene Verfahren erstellt werden. Die erste Bestimmung der Breite, ist Bestimmt durch die Differenz von Minimum und Maximum dividiert durch eine fixe Anzahl an Wertebereichen (Bestimmung der Anzahl der Klassen durch Erfahrungswerte oder per Faustregel siehe Tabelle 3.1), als zweites Anhand der Sturges Regel $k = 1 + \log_2 n = 1 + 3,3 \cdot \log_{10} n$ und zuletzt durch Scott [SCO79], welche die Streuung der Daten beachtet und den Interquartilsabstand zur Hilfe nimmt. In Abbildung 3.2 sieht man unterschiedliche Histogramme. Diese Histogramme basieren auf der gleichen zeitbehafeten Messreihe, bei unterschiedlicher Breite der Klassen.

| Anzahl der Werte | Klassen |
|------------------|---------|
| <50 | 8 |
| 50 bis 100 | 16 |
| 100 bis 250 | 16 |
| >250 | 32 |

Tabelle 3.1: Faustregel für Histogramme

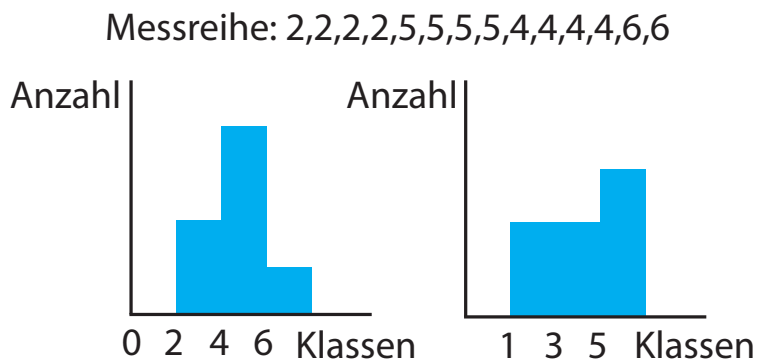


Abbildung 3.2: Unterschiedliche Histogramme mit gleichem Datensatz

3 Related Work

Die kumulative Summe (CUMSUM) beschreibt eine Methode zur Erkennung von Änderungen von Parametern wenn ein Eingriff nötig wird. Als grafische Visualisierung lässt sich für die CUMSUM das V-Mask Diagramm benutzen. Die CUMSUM beschreibt hierbei die Differenzen von Datenwerten x_n und Wahrscheinlichkeitsfunktionen w_n (Bsp. Likelihood). Die Summe ist als $S_0 = 0, S_n = \max(0, S_{n-1} + x_n + w_n)$ definiert und erkennt Abweichungen negativer und positiver Art in bestimmten kritischen Bereichen.

3.1.4 Normalverteilungen

Die Abweichung von naturellen, wirtschaftlichen und anderen Messwerten beruht meist auf dem zentralen Grenzwertsatz, der eine Verteilung von genügend unabhängigen Zahlen als Normalverteilt ansieht. Diese Verteilung zieht man für die Erkennung von Messfehlern und Abweichungen herbei. Der Hochpunkt bzw. Peak der Normalverteilung ist in den meisten Fällen der Mittelwert und Durchschnitt aller Messpunkte. Für die Abweichungen und Verteilungen von Messpunkten um diesen Mittelwert wird die Varianz σ benötigt und folgende Aussagen lassen sich über die Messpunkte bei einer Normalverteilung sagen:

- 50% aller Werte haben eine Abweichung von höchstens $0,675\sigma$ vom Mittelwert
- 90% aller Werte haben eine Abweichung von höchstens $1,645\sigma$ vom Mittelwert
- 95% aller Werte haben eine Abweichung von höchstens $1,960\sigma$ vom Mittelwert
- 99% aller Werte haben eine Abweichung von höchstens $2,576\sigma$ vom Mittelwert

Die mathematische Definition der Normalverteilung ist durch die Definition 3.2 und der z-Verteilungen 3.3 gegeben. Die Standardnormalverteilung wird mit den Verteilungseigenschaften mit $\mu = 0$ und $\sigma^2 = 1$ definiert.

$$F(x) = \frac{1}{\hat{s}\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\bar{x}_{average}}{\hat{s}}\right)^2} dt \quad (3.2)$$

$$z(\alpha) = z(1 - \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt \quad (3.3)$$

3.2 Data-Mining Verfahren

Neben statistischen Verfahren spielen in der Analyse von Prozessen die Verfahren aus dem Data-Mining eine große Rolle. Diese können für die Verknüpfung und Generierung von Wissen und für die Analyse von wichtigen Bestandteil sein. Zudem können durch unterschiedliche Verfahren, wie das Clustering oder die Regression, statistischen Verfahren in ihrer Genauigkeit verbessert und beschleunigt werden.

3.2.1 Clusteranalyse

Bei einer Clusteranalyse beschreibt man ein Verfahren, das Gleichheiten und Ähnlichkeiten von Eigenschaften erkennt. Diese gefundenen Ähnlichkeiten nennt man Cluster. Hierbei können Cluster nicht nur eine Dimension eines Attributes beherbergen, sondern auch mehrere Dimensionen. Diese Analyse der Cluster lassen sich in den meisten Fällen erst durch eine nachträgliche Analyse identifizieren, um weiteres Wissen über die Daten zu erzeugen. In Abbildung 3.3 wird ein zeitlicher Ablauf mehrerer Instanzen gezeigt. Ähnliche zeitliche Abläufe lassen sich zusammenfügen und den Durchschnitt, wie Abbildung 3.3 zu sehen berechnen. Weiter lässt sich aus vielen Messreihen ein durchschnittlicher Fall bestimmen (siehe Abbildung 3.3, P_4). Typische Cluster-Eigenschaften in Geschäftsprozessen sind die zeitliche Ablaufreihenfolge von Aktivitäten einer Instanz und gleicher Aktivitäten in einem Prozess.

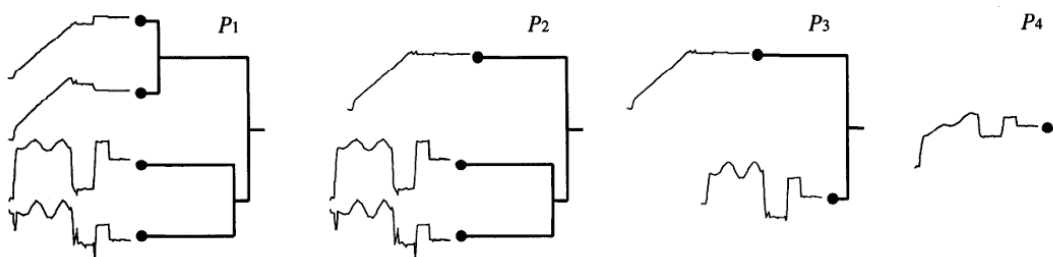


Abbildung 3.3: Clustering von Messreihen [KP98]

3.2.2 Regression

Die Regression beschreibt eine mathematische Methode, um an Messdaten mit Parametern mit möglichst wenig Widersprüchen und Abweichungen an eine bestmögliche Funktion anzunähern. In den meisten Fällen werden Messdaten als eine angenäherte Kurve dargestellt und die Abweichung zu den realen Messdaten bestimmt. Hierbei existieren einige Verfahren wie die kleinsten Quadrate, stochastische, funktionale Modelle, das Gaußesche Eliminationsverfahren und selbst graphische Verfahren für den Benutzer. In einer Instanz können die Anfangs- und Endzeitpunkte der Aktivitäten durch eine Regression dargestellt und durch die Clusteranalyse verfeinert werden.

3.3 Analysen auf Prozesse und Prozesselementen

Eine Analyse für Prozesselementen basiert in den meisten Fällen auf den oben beschriebenen Verfahren. Jedoch muss für jedes statistische oder Data-Mining Verfahren der Prozess aufbereitet und für die Analyse angepasst werden. Hierbei spielen die zeitbehafteten Attribute eines Prozesses und dessen Instanz eine große Rolle. In diesem Abschnitt werden Verfahren vorgestellt, die Prozesse analysieren und für das Monitoring bzw. die Vorhersagen Daten bereitstellen.

3.3.1 Verteilungen einzelner Dauern von Aktivitäten

Die Dauern von einer Aktivität und Abstände zwischen Aktivitäten lassen sich als Messpunkt in einer zeitbehafteten Datenreihe sehen. Auf dieser Datenreihe lässt sich ab einem gewissen Maß an Messpunkten die Verteilungsfunktion der jeweiligen Reihe bestimmen[RSW12]. Bei einer Verteilung kann man davon ausgehen, dass die Werte mit der Wahrscheinlichkeit unter und über 1% als Ausreißer gesehen werden können. Zudem können bei Entscheidungen in Prozessen mehrere Peaks und daher mehrere Normalverteilungen in einer Messreihe entstehen (siehe Abbildung 3.4). Dadurch können schnell kritische Bereiche und Ausreißer erkannt werden. Diese Grenzen unterhalb des 1% unterteilt die Messreihe in Intervalle, die durch einen kritischen und einen korrekten

3.3 Analysen auf Prozesse und Prozesselementen

Bereich beschrieben werden. Der kritische wird durch die untere bzw. obere Grenze eingegrenzt und der korrekte Bereich durch die Werte zwischen der unteren/oberen Grenze, bei mehreren Verteilungen in einer Messreihe.

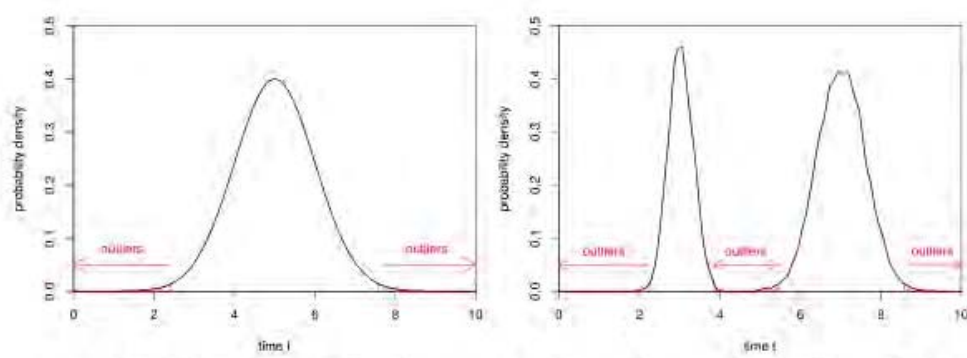


Abbildung 3.4: Verteilungen auf zeitbehafteten Messreihen [RSW12]

3.3.2 Filter und Statistische Verfahren

Neben den Verteilungen spielen einfache statistische Verfahren für die Vorhersage eine große Rolle. Berechnungen des Mittels, Medians, Minimum, Maximum und der Varianz gehören zum Standard. Diese Verfahren wurden in Abbildung 3.5 angewandt und an den Prozess annotiert. Das besondere an den Herangehensweisen an dieser Methode, ist das Benutzen von Filtern und Berechnung des Fehlers gegenüber des Mittels. Zudem können auf Prozesse und seinen Daten-Logs verschiedene Sichten und Zusammenfassungen dessen Ereignissen erstellt werden [ASS11]. Diese Verfahren sind für die grundlegende Transformation und Analyse notwendig und bieten eine breite Grundlage für weitere Analysen auf Messreihen. Weiter können Sichten mit verschiedenen Berechnungsgrundlagen zum Instanzenende oder Instanzanfang genutzt werden.

3 Related Work

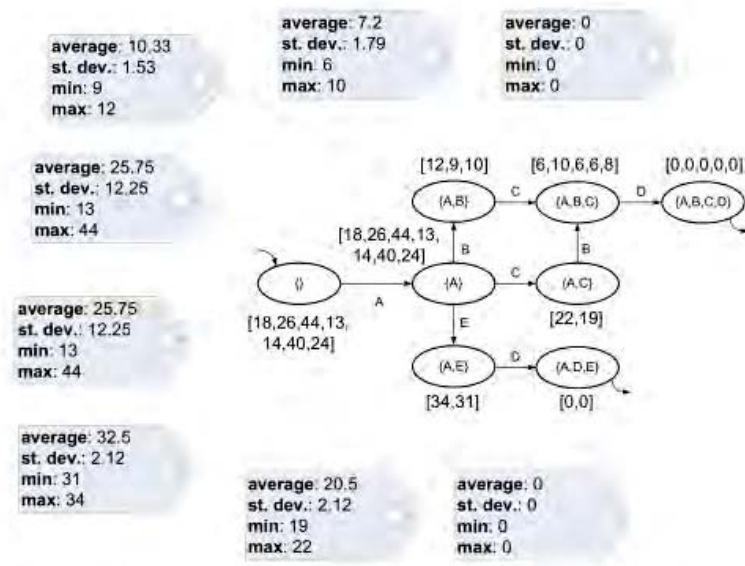


Abbildung 3.5: Einfache Berechnung und Analysen der Zeitaspekte [ASS11]

3.4 Monitoring von Prozessen

Die Analyse der Daten stellt die Grundlage für das Monitoring dar. Diese Analysen generieren aus verschiedenen Ereignis-Logs Informationen über einen Prozess, welche für den Monitor verwendet werden. Hierbei spielt die Abhängigkeit zwischen Aktivitäten, Instanzen, den parallel laufenden Instanzen und die Umwandlung des Geschäftsumfeld eine große Rolle beim Monitoring. In diesem Abschnitt werden einige Verfahren für das Zusammenspiel von Ereignissen, Instanzen und weiteren Monitoringelementen vorgestellt.

3.4.1 Konversation zwischen Instanzen und Events

Um ein Monitoring zu ermöglichen, werden nach dem Erzeugungs-Zeitstempel die Ereignisse geordnet. Dadurch können Schlüsse auf die Instanzen gezogen und eine Umwandlung in neues Wissen generiert werden. Diese Umwandlung und Abhängigkeiten der Instanzen, Ereignissen und die Kommunikation ist in Abbildung 3.6 gezeigt.

Eine Änderung in der einen Komponente erzeugt eine Änderung in den Abhängigkeiten. In der einen Komponente werden bei der Ausführung von Instanzen Ereignisse in den Log geschrieben und zur Überwachung der laufenden Instanz zur Hilfe genommen. Dabei ist die Umwandlung und Bereitstellung der Informationen aus dem Log für die Überwachung essentiell. Diese Umwandlung kommuniziert mit den Instanzen und stellt für die Ausführung der Instanzen Informationen über die Ereignisse bereit [BDDW07]. Diese beiden Komponenten interagieren durchgehend miteinander, um eine Vorhersage über Komponenten der Instanz treffen zu können. Dieser Aufbau zeigt den grundlegenden Aufbau eines Monitors. Durch eine Extraktion der Daten aus den Logs, der Transformation der Daten, das Analysieren der Daten und im letzten Schritt das Laden der Daten in den Monitor beschreibt den ersten Schritt der Überwachung von Prozessen. Nach diesem Laden werden die Ereignisse mit den Analyse-Daten verglichen.

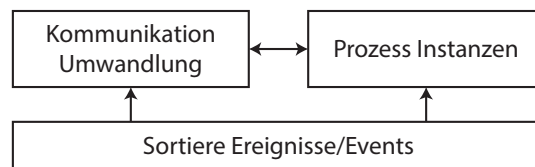


Abbildung 3.6: Zusammenspiel der Komponenten [BDDW07]

3.4.2 Definierte Überwachungspunkte in Prozesse

Wie in den Grundlagen beschrieben, wird eine Aktivität in einem Geschäftsprozess durch verschiedene Ereignisse (engl. Events) dargestellt. Diese Aktivität kann mehrere Zustände, wie in Kapitel 2 beschrieben annehmen. Diese Änderung kann als Event gesehen und als Überwachungspunkt deklariert werden [HKRS12]. Wie in Abbildung 3.7 zu sehen, wurde bei dem Überwachungspunkt M_1 , das Event des Beginns ausgewählt und bei den Überwachungspunkten M_2 und M_3 das Event des Terminierens der Aktivität. Weiter werden Systeme und verschiedene Verfahren festgelegt, in welchen die Ereignisse behandelt werden. Hierbei kann die Art des Monitorings festgelegt werden

3 Related Work

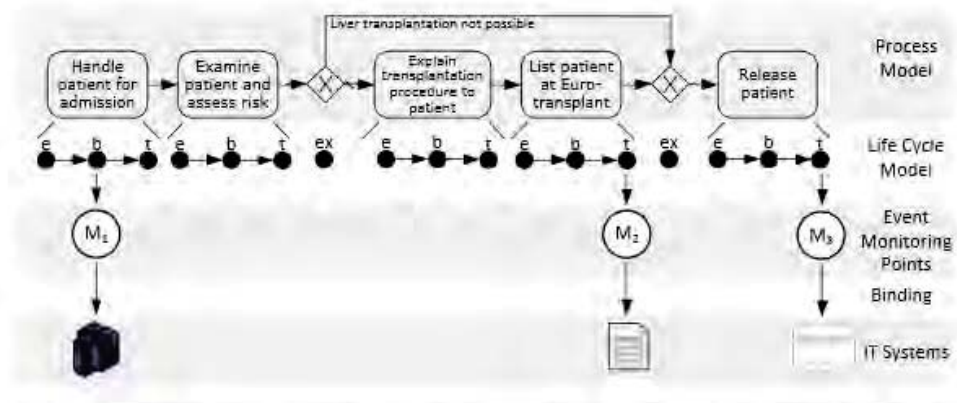


Abbildung 3.7: Prozess in Verbindung mit Events [HKRS12]

3.4.3 Korrelationen von Ereignissen

Neben einer Einteilung der Ereignisse in die Zustandsänderung einer Aktivität können Ereignisse je nach ihrer Eigenschaft und Ort der Entstehung eingestuft werden [BDDW07]. Hierbei wird zwischen *Action* und *Communication* Events unterschieden. Action Events werden von einem Benutzer ausgeführt und können Communication Events auslösen. Diese Communication Events stoßen bei einem anderen Benutzer ein *Communication Event* an, welches das *Action Event* des Benutzers auslöst. Diese werden daher durch die Abhängigkeit der Zeit, des Benutzers, der Ort der Entstehung und die Art des Events definiert. Die Abhängigkeiten der Events mit den Benutzern ist in Abbildung 3.8 zu sehen, dabei interagieren die Benutzer mit den Action Events, welche in Abhängigkeit mit den Communication Events stehen. Diese Events und Abhängigkeit können für das Monitoring in verschiedene Patterns eingeteilt werden die nach [BDDW07] folgend definiert sind: 1. Korrelation, definiert einen Mechanismus, welcher eine Abfolge von Events direkt oder transitive in Verbindung mit Zeitstempel und zeitbehafteten Intervallen darstellt, 2. Konversations, beschreiben eine Abhängigkeit zwischen zwei Konversationen in verschiedenen Instanzen, 3. eine Instanz in Verbindung mit Abhängigkeiten, beschrieben durch eine Gruppierung von Events in einer Instanz in Kombination mit den Abhängigkeiten eines Benutzer. Diese Definitionen stellen Gruppierungen dar, die das Monitoring auf verschiedene Bereiche einschränkt.

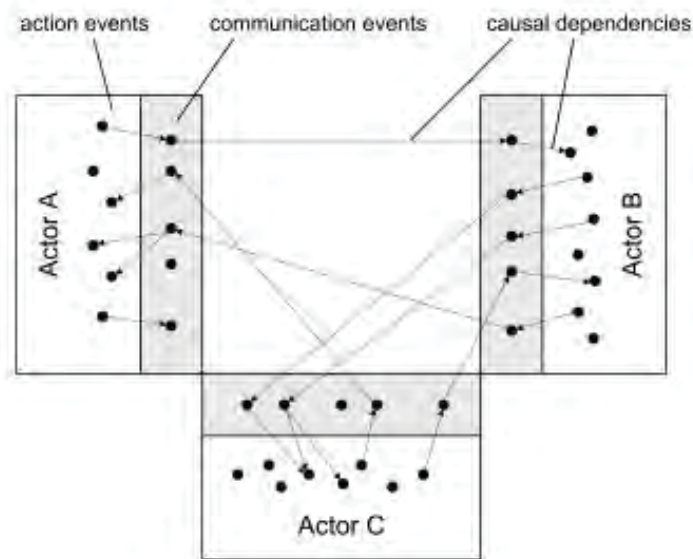


Abbildung 3.8: Verschiedene Event-Typen [BDDW07]

3.4.4 Berechnen der Intervalle zum Schluss

Abstände zu überwachen ist das Essentielle des Monitorings. Dabei bieten sich zwei Möglichkeiten den Fortschritt einer Instanz zu bestimmen. Zum einen im Verhältnis zum Anfang der Instanz und zum anderen bis zum Ende. Es wird geschaut ob die Intervalle valide oder nicht valide sind [CP99] und es wird zwischen strikten (siehe Abbildung 3.9) und längsten Intervallen unterschieden. Das strikte Intervall beschreibt die Zeit die eingehalten werden muss und das längste was im schlimmsten Fall eintreten kann. Diese Abstände werden hierbei zum Ende der Instanz für die Vorhersage gemessen.

Neben den Monitoringverfahren im Allgemeinen, spielen in dieser Arbeit die Darstellung von Zeiten in Prozessen eine Rolle und werden im folgenden Abschnitt beschrieben.

3 Related Work

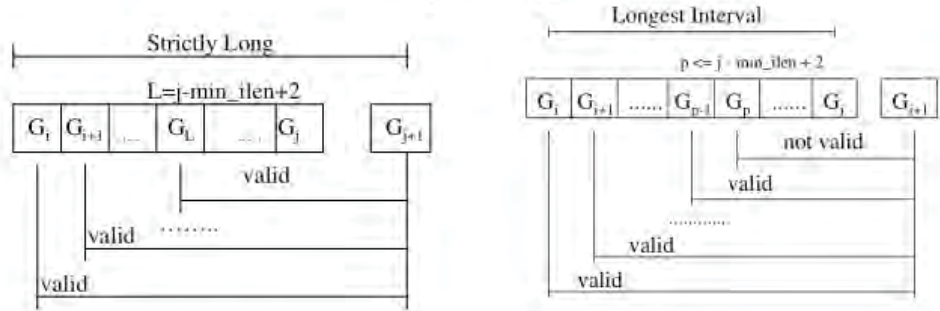


Abbildung 3.9: Abstände zum Schluss [CP99]

3.5 Zeitmanagement in Prozessen

Neben den Zeitaspekten aus Abschnitt 2.2, existieren noch weitere Verfahren für die Darstellung und Transformation von zeitbehafteten Geschäftsprozessen. Die Zeitaspekte aus den Grundlagen stellen nur die nötigsten Voraussetzungen für diese Arbeit dar. Jedoch haben in der Vergangenheit unterschiedliche und komplexe Verfahren aus der Künstlichen Intelligenz und verschiedenen Planern Einzug in die Welt der Prozess-Management-Systeme gehalten.

3.5.1 Transformation eines zeitbehafteten Prozessgraphen

Ein typischer Prozessgraph aus dem *ADEPT2*-Prozess-Metamodell lässt sich für die Weiterverwendung für Algorithmen und Darstellungs-Möglichkeiten kaum verwenden, da diese Sicht des Prozesses für den Benutzer optimiert ist. Daher muss der Prozessgraph transformiert und angepasst werden. In dieser Anpassung, werden einzelnen Aktivitäten in zwei separate Knoten aufgesplattet. Gateways, die als Raute dargestellt sind, erhalten einen separaten Knoten im neuen Prozessgraphen. Dies ist in Abbildung 3.10 zu sehen. Hierbei wird die Dauer einer Aktivität gleich behandelt wie ein Abstand zwischen zwei Aktivitäten. Dieser neue Abstand hat dabei die zeitlichen Eigenschaften von einer minimalen und maximalen Dauer. Daher wird zwischen einem Prozessmodell und dem Zeitmodell unterschieden [LR14]. Das Zeitmodell kann für das Prozessmodell je nach

Transformationseigenschaften unterschiedlich sein und für die jeweilige Algorithmen der weiteren zeitbasierten Berechnungen genutzt werden.

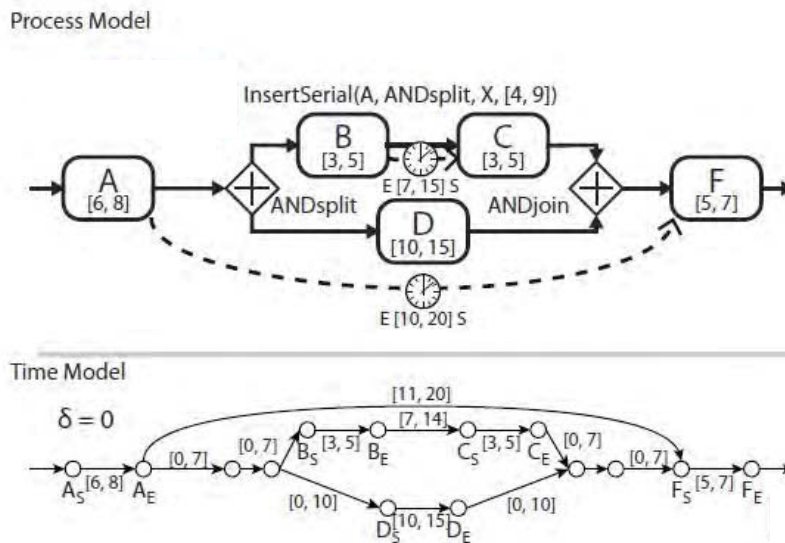


Abbildung 3.10: Transformation von zeitbehafteten Prozessen [LR14]

3.5.2 Transformation eines Prozessgraphen in Anlehnung an CSTNU

In diesem Verfahren, werden Prozesselemente durch weitere Zeitaspekte aus der Tabelle 2.3 erweitert. Hierbei wurde im Vergleich zum vorigen Abschnitt die Prozesselemente noch weiter verfeinert und in drei anstatt zwei Knoten aufgespaltet [LPCR13]. Dadurch lassen sich weitere zeitbasierte Elemente, wie die minimale Dauer, die maximale Dauer im besten Fall und die maximale Dauer im schlechtesten Fall darstellen und für die weitere Berechnung anwenden lassen. Dies ist beispielsweise in Abbildung 3.11 für die Aktivität A zu sehen. Durch diese Anpassungen des Zeitmodells lassen sich alle Eigenschaften der Zeitaspekte aus Tabelle 2.3 berechnen und aus einem Prozessmodell extrahieren.

Nach den Grundlagen und den verwandten Themengebieten, werden die Verfahren bewertet und ein Fazit gezogen, welche für die Darstellungs- und Analyse-Möglichkeiten interessant sind. Diese kurze Bewertung wird im nächsten Abschnitt vorgestellt.

3 Related Work

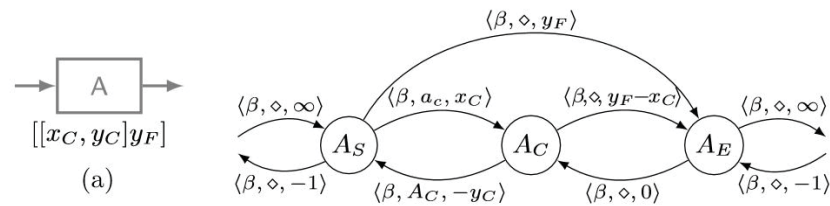


Abbildung 3.11: Aktivität mit Mapping auf CSTNU [LPCR13]

3.6 Fazit

Nach der Betrachtung verschiedener statistischer und Data-Mining Verfahren, wurde in diesem Kapitel die Analyse von zeitbasierten Prozesselementen, Verfahren auf zeitliche Messreihen und verschiedene Transformationen vorgestellt. Hierbei haben sich für die Analyse von zeitbehaftete Geschäftsprozesse einige Verfahren alleine und in Kombination miteinander von Vorteil erwiesen. Im nächsten Kapitel werden Verfahren der Boxplots und Histogramme für die Analyse und Darstellung herangezogen. Diese Verfahren können mit den meisten Verteilungen umgehen und beherrschen zum Teil eine automatisierte Eliminierung der Ausreißer in den Messreihen. Weiter wurde aufbauend auf den Abschnitten 3.4 & 3.5 einige Verfahren für das Monitoring und der Vorhersage vorgestellt, welche im Kapitel 5 in abgewandelter Weise verwendet werden.

4

Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

Im Folgenden werden die in Kapitel 1 diskutierten Anforderungen an die Darstellung und Analyse von zeitbehafteten Geschäftsprozessen umgesetzt. Eine solche Darstellungs- und Analyse-Möglichkeit nimmt bei der Optimierung einzelner Geschäftsprozesse eine wichtige Rolle ein. Beispielsweise können durch die Verknüpfung von historischen Prozess-Event-Logs und den Informationen des Prozess-Management-Systems neue Informationen über den Geschäftsprozess gewonnen werden. Da diese Prozess-Event-Logs den Erstellungs-Zeitstempel der Ereignisse beinhalten, können die Zeitstempel für die Generierung von Aussagen über zeitlichen Eigenschaften von Instanzen und Aktivitäten genutzt werden. Diese Aussagen über die zeitlichen Eigenschaften bewerten Prozesselemente und ermöglichen es, Schlüsse über den Zustand der Prozesselemente zu ziehen. Um dieses Verfahren der Analyse und Aufbereitung von zeitbehafteten Daten

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

zu ermöglichen, können diese durch einen ETL-Prozess Referenz analysiert und für die spätere Darstellung in einer geeigneter Form aufbereitet werden. Dieser ETL-Prozess ist in mehrere Schritte unterteilt: der erste Schritt ist die Daten Extraktion, der zweite die Transformation und zuletzt noch das anpassen der Form für das Laden in folgende Systeme. Dieser ETL-Prozess sollte die Darstellung der zeitbehafteten Daten der in Kapitel 1 diskutierten Anforderungen, erfüllen. Diese Anforderungen beinhalten die typischen Prozesselemente, wie Sequenzen, Parallelität, Entscheidungen und Schleifen. Für diesen ETL-Prozess und die damit einhergehenden Probleme werden im Folgenden Lösungsansätze aus der Statistik präsentiert, erörtert und anhand den Anforderungen bewertet.

Im ersten Abschnitt, wird der ETL-Prozess beschrieben, welcher die Analyse der zeitbehafteten Daten beinhaltet. Diese Analyse der zeitbehafteten Daten wird in den folgenden zwei Abschnitten beschrieben. Zum einen die Analyse- und Darstellungs-Möglichkeit mit Histogrammen und zum anderen mit den Boxplots. Diese beiden Visualisierungen von zeitbehafteten Messreihen werden an einem *ADEPT2*-Prozess annotiert und behandelt. Im vierten Abschnitt wird die Analyse- und Darstellungs-Möglichkeit um die Transformation des Prozessgraphen erweitert. Danach wird eine Möglichkeit zur Beschreibung von Organisationsstrukturen, welche in den Prozesselementen eine Rolle spielen, beschrieben. Abgeschlossen wird das Kapitel durch einen kurzen Vergleich, welcher auf die wichtigsten Punkte und Probleme bei der Darstellung und Analyse eingeht.

4.1 ETL-Prozess

Für die Gewinnung der Analyse-Ergebnisse aus den Event-Logs der zeitbehafteten Geschäftsprozesse müssen die Informationen und Daten aus den verschiedenen Quellen herangezogen und verarbeitet werden. Bei dieser Verarbeitung müssen die Daten aus den unterschiedlichen Quellen so angepasst und bearbeitet werden, dass diese für die Darstellung im PMS und im Monitor verwendet werden können. Diese Verarbeitung wird mit Hilfe eines ETL-Prozesses (Extract, Transform, Load) durchgeführt [BSS00]. Dieser Prozess ist ein drei Schritte aufgeteilt (siehe Abbildung 4.1). Der erste Schritt besteht aus dem **Extrahieren** der Informationen und der Daten aus den verschiedenen Quellen.

Diese Daten werden dadurch in einen zentralen Arbeitsbereich des ETL-Prozesses bereitgestellt. Auf diesem Arbeitsbereich arbeitet der zweite Schritt die **Transformation**. Diese Transformation erkennt die passende Datentypen der Quelldaten und verknüpft die verschiedenen Quellen. Im selben Schritt werden die in den Abschnitten 4.2, 4.3 & 4.4 vorgestellten Analyseverfahren angewandt und ausgeführt. Dabei werden fehlerhafte Datenreihen und Datenpunkte entfernt. Im letzten Schritt des **Loading** werden die Daten in ein passendes Format für die Darstellung im PMS und des Monitors gebracht und anschließend in diese geladen.

4.1.1 Extraktion

Der erste Schritt des ETL-Prozesses besteht darin die verschiedenen Informationen aus den vorliegenden Quellen zu extrahieren und im Arbeitsbereich des ETL-Prozesses bereitzustellen. Die benötigten Informationen liegen dabei in den unterschiedlichen Quellen in verschiedenen Formaten und Strukturen vor und müssen daher zunächst für den Arbeitsbereich aufbereitet werden. Diese Aufbereitung geschieht durch speziell angepassten Verfahren für die unterschiedlichen Datenquellen. Dabei spielt die Art der Verbindung zu der Quelle, die Struktur der Informationen und die Verteiltheit des Quellsystems für die Anpassung und Bereitstellung eine wichtige Rolle.

Diese unterschiedlichen Datenquellen müssen in verschiedenen Zyklen aktualisiert werden. Änderungen an einem Prozess müssen versioniert und die Extraktion teils neu angewandt werden. Nach einem gewissen periodischen Zyklus müssen die neu generierten Daten des Prozess-Event-Logs dem Arbeitsbereich hinzugefügt werden. Dadurch wird gewährleistet, dass die Analyse auf den aktuellen Daten arbeitet und ein korrektes Ergebnis liefert. Zudem werden bei der Extraktion fehlerhafte und nicht komplette Daten beim Bereitstellen in den Arbeitsbereich entfernt. Die Art, wie die Daten aus den unterschiedlichen Quellen im Arbeitsbereich bereitgestellt werden, ist je nach Verfahren unterschiedlich. Die Anpassung auf das passende Format im Arbeitsbereich wird im Schritt der Transformation erledigt.

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

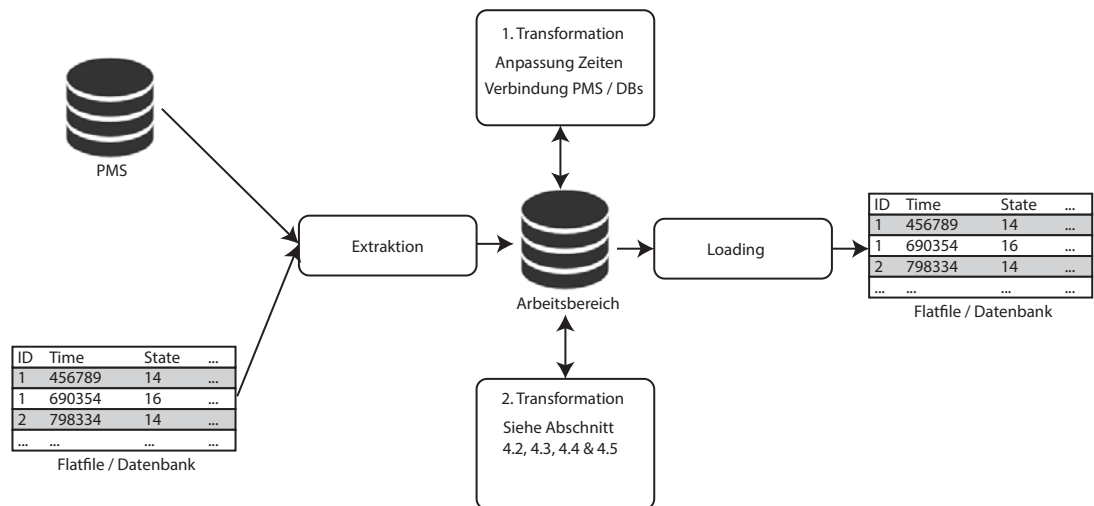


Abbildung 4.1: ETL-Prozess

4.1.2 Transformation

Nach dem Bereitstellen und Extrahieren der Daten aus den verschiedenen Quellen in den Arbeitsbereich müssen diese Informationen für den Analyse-Prozess angepasst und transformiert werden. Diese Transformation der Daten läuft in zwei Schritten ab: Im ersten Schritt werden die Prozess-Event-Logs in eine tabellenartige Struktur gebracht und die Datentypen erkannt. Um die Informationen eines PMSs zu erhalten und die Prozessstrukturen zu bestimmen, wird aus dem Prozessgraphen eine tabellenförmige Struktur erstellt. Im gleichen Schritt wird die Organisationsstruktur der Geschäftsprozesse bestimmt und mit den anderen transformierten Daten im Arbeitsbereich verbunden. Die Zeitstempel der einzelnen Einträge werden im Arbeitsbereich im Verhältnis zu ihrem Instanz-Start umgeformt und angepasst. In Abbildung 4.1 werden beispielhafte Quellen gezeigt und die Reihenfolge des ETL-Prozesses dargestellt. Im zweiten Schritt der Transformation werden die verknüpften Daten im Arbeitsbereich analysiert. Diese Analyse-Möglichkeiten werden in den Abschnitten 4.2, 4.3 & 4.4 erläutert.

4.1.3 Loading

Der letzte Schritt des ETL-Prozesses beschreibt das Laden der Informationen und Daten aus dem Arbeitsbereich in ein passendes Format für die entsprechende Anwendung. Hierbei müssen die analysierten Daten durch verschiedene Schemata für das Format des PMS oder des Monitors angepasst werden. Je nach Verfahren liefert das *Loading* verschiedene Formate zurück, die durch das gewählte Analyseverfahren festgelegt werden. Diese Daten liefern zum einen die Informationen für das PMS, welches dadurch die Annotationen an die Prozesselemente anbringen kann und zum anderen die Bewertungskriterien der Prozesselemente für den Monitor. Diese Darstellungs- und Analyse-Möglichkeiten und die passenden Formate werden im Folgenden beschrieben.

4.2 Zeitbasierte Histogramme

Nach der Extraktion der Daten aus den verschiedenen Quellen in den Arbeitsbereich und dem ersten Schritt der Transformation der Daten werden diese Daten im zweiten Schritt der Transformation verarbeitet und analysiert. Hierbei stellt sich die Frage, wie zeitbehaftete Daten für einen Mensch gut lesbar dargestellt werden können und wie der Benutzer eine Schlussfolgerung auf die zeitlichen Eigenschaften von Prozesselementen ziehen kann. Diese Anforderungen an die Lesbarkeit und Wissensgewinnung, welche die Verfahren erfüllen sollen, wurden in Kapitel 1 beschrieben. In diesem Abschnitt wird die Analyse-Möglichkeit der zeitbasierte Histogramme (siehe Kapitel 3) im zweiten Schritt des ETL-Prozesses gewählt und erläutert.

4.2.1 Erstellung von zeitbasierten Histogrammen

Die Analyse des Verfahrens der zeitbasierten Histogramme findet im zweiten Schritt des ETL-Prozesses statt. Nach der Aufbereitung der Zeitstempel im Verhältnis zum Instanzanfang und der Verknüpfung von dem angepassten Prozess-Event-Log mit den Information aus dem PMS wird der zweite Schritt der Transformation durchgeführt. Bei

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

dieser Transformation werden zunächst für jede Aktivität der unterschiedlichen Prozesse die Dauer und anschließend der Abstand zwischen aufeinanderfolgende Aktivitäten berechnet. Diese Dauer einer Aktivität und der Abstand von aufeinanderfolgende Aktivitäten, wird als unabhängige Datenreihe gesehen und kann deshalb für die Transformation der zeitbasierten Histogrammen verwendet werden. Bei diesen Histogrammen werden die Dauern einer Aktivität und die Abstände von aufeinanderfolgende Aktivitäten in Klassen bestimmter Breite eingeteilt. Die Klassen stellen hierbei eine Intervall dar, in welchen Messpunkte der Dauer einer Aktivität oder Abstände von aufeinanderfolgende Aktivitäten einsortiert werden. Zuletzt wird die Anzahl der Messpunkte gezählt und somit ein Histogramm erstellt. Die Breite der Klasse wird durch die Differenz des maximalen und dem minimalen Wertes durch die gewünschte Anzahl an Klassen geteilt. Daher ergibt sich die Breite der Klassen auf einer Messreihe $a[]$ mit n Anzahl an Klassen, $(\max(a[]) - \min(a[]))/2^n$. In Abbildung 4.2 wird für die Aktivität A, mit den aus dem ersten Schritt ermittelten Dauern das zeitbasierte Histogramm erstellt. Das Minimum und Maximum wird bestimmt und durch den Wert der Faustregel aus Tabelle 4.1 geteilt. Im nächsten Schritt werden die Dauern der Aktivität in die Klassen einsortiert und die Anzahl der vorhanden Werte in der Klasse ausgewertet. Das resultierende Histogramm ist in Abbildung 4.2 zu sehen.

| Anzahl der Werte | Klassen | n |
|------------------|---------|---|
| <50 | 8 | 3 |
| 50 bis 250 | 16 | 4 |
| >250 | 32 | 5 |

Tabelle 4.1: Faustregel für die Klassen

Nach der Erstellung des Histogramms wird versucht die Anzahl der Klassen zu verringern, um eine bessere Lesbarkeit und Übersichtlichkeit zu erreichen. Zudem lassen sich durch diesen Schritt die Aussagen, neben dem Maxima und Minima, über die Verteilung, den Modus und den ungefähren Median leichter treffen. Diese Zusammenführung von Klassen ist jedoch nur möglich, wenn nebeneinander liegende Klassen geringe Unterschiede aufweisen (siehe Abbildung 4.3). Da Histogramme nicht ausreißerneutral sind und bei Aktivitäten und deren Abständen bei fehlerhaften Ausführungen einen extremen

4.2 Zeitbasierte Histogramme

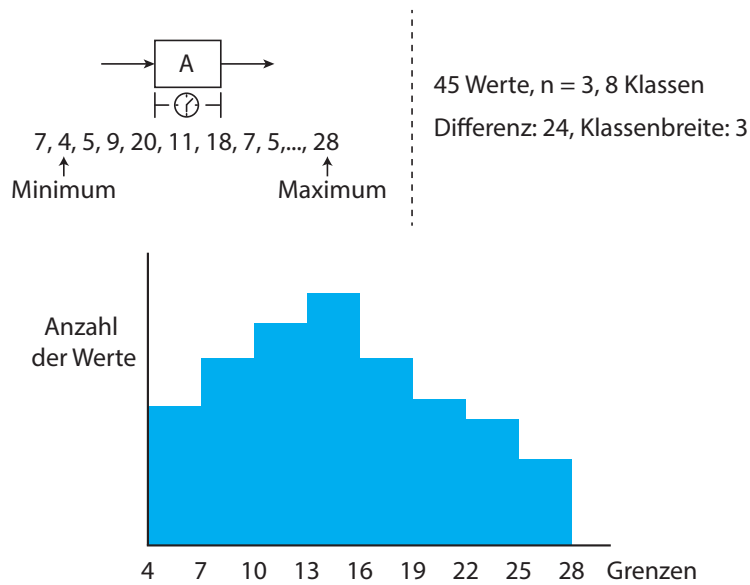


Abbildung 4.2: Erstellung des zeitbasierten Histogramms

Wert annehmen können, wird im nächsten Abschnitt die Eliminierung von Ausreißern besprochen.

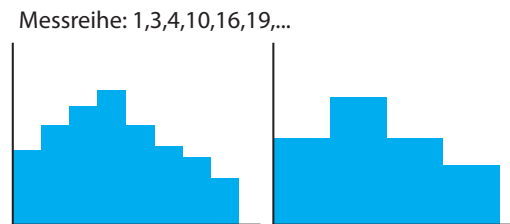


Abbildung 4.3: Zusammenführen von Klassen

4.2.2 Ausreißer in Histogrammen

In einer Instanz können zu lange oder zu kurze Ausführungszeiten von Aktivitäten oder ein zu langes Warten bis zum Beginn der Aktivität entstehen. Solange diese Abweichungen nicht zu groß werden, werden diese Werte im Histogramm ohne Pro-

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

blematik dargestellt und es entsteht ein Peak in der Mitte des Histogramms. Existieren aber extreme Ausreißer gibt es ein Problem mit der Größe der Klassen, da sie durch $(\max(a[]) - \min(a[]))/2^n$ definiert ist. Das heißt, bei einem großen Maximum, wie in Abbildung 4.4 zu sehen wird eine gefüllte Klasse beim Minimum, eine belegte Klasse beim Maximum und viele nicht gefüllte Klassen entstehen. Dadurch ist der Wissensgewinn für den Benutzer sehr gering und die Aussage über die Verteilung ist gleich null. Neben Minimum und Maximum lassen sich keine Schlüsse über die Ausführungen der Aktivität oder den Abständen von aufeinanderfolgenden Aktivitäten treffen. Deswegen müssen zunächst die extremen Ausreißer bei dem Verfahren der zeitbasierten Histogrammen aus den Messreihen entfernt werden. Dies wird im folgenden Abschnitt beschrieben.

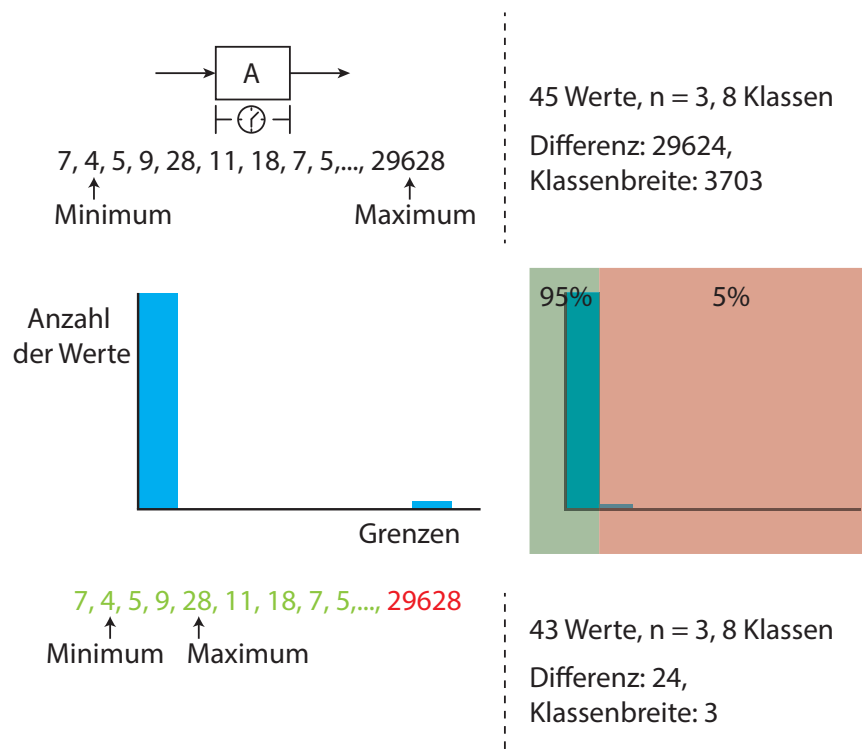


Abbildung 4.4: Ausreißerproblem bei Histogrammen mit Eliminierung

4.2.3 Angepasste zeitbasierte Histogramme

Ohne eine Anpassung und Entfernung der extremen Ausreißer in einer Messreihe kann man aus Histogramme kaum Informationen gewinnen. Deshalb müssen sie durch geeignete Verfahren entfernt werden. Hierbei hat sich das Entfernen von gering gefüllten Klassen bewährt. Bei dieser Entfernung werden die Klassen nach ihrer Größe absteigend sortiert. Da bei einer normalverteilten Messreihe ca. 95% der Werte in einem Intervall um den häufigsten Wert liegen, können 5% der Messwerte, der kleinsten Klassen entfernt und eliminiert werden. Bei dieser Eliminierung dürfen nur komplette Klassen entfernt werden. In Abbildung 4.4 sieht man die Sortierung der Klassen und den extremen Ausreißer. Diese Klasse, welche den Ausreißer beinhaltet wird entfernt und somit auch der Ausreißer aus der Messreihe getilgt. Nach dieser Entfernung des Ausreißers wird das Verfahren der Eliminierung von Ausreißern in zeitbasierten Histogrammen sooft wiederholt, bis keine Klassen mehr entfernt werden müssen. Im folgenden Abschnitt wird die Annotation der zeitbasierten Histogramme an den Prozessen gezeigt.

4.2.4 Darstellung der Histogramme in Prozessen

Nach dem ETL-Prozess können die Daten der Histogramme in das PMS geladen werden. Das PMS nutzt hierbei die Größe und die Höhe der jeweiligen Klassen des Histogramms. Die Histogramme werden an die jeweilige Aktivität als Dauer und an die Kontrollflusskannte der aufeinanderfolgenden Aktivitäten annotiert. Diese Annotationen sind in Abbildung 4.5 zu sehen. Hierbei lassen sich über die Verteilungen der Messreihen von Dauern und Abständen zwischen Aktivitäten Informationen über das Minimum, Maximum und den ungefähren Durchschnitt ablesen und bestimmen. Die Information der Verteilung lässt sich ablesen, wie am Abstand von Aktivität A zur Aktivität B der harmonischen Verteilung In Abbildung 4.5 zu sehen ist. Außerdem zu sehen ist zusätzlich, dass die Dauer von A einer harmonischen Verteilung ähnelt, jedoch einen kleinen Einbruch in der sechsten Klasse besitzt. Dagegen ist das Histogramm von Aktivität D nach rechts verschoben. Daher dauert diese Aktivität in den meisten Fällen länger und stellt eine untypische harmonische Verteilung dar. Neben den Bewertungen von Dauern von Aktivitäten und Abstände von aufeinanderfolgenden Aktivitäten können im Prozess selbst

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

noch Abstände für die Analyse erstellt werden. Diese können wie in den Grundlagen (siehe Kapitel 2.2) von den jeweiligen Start- bzw. Endpunkten der ersten Aktivitäten zu den Start- bzw. Endpunkten der zweiten Aktivität bestimmt werden.

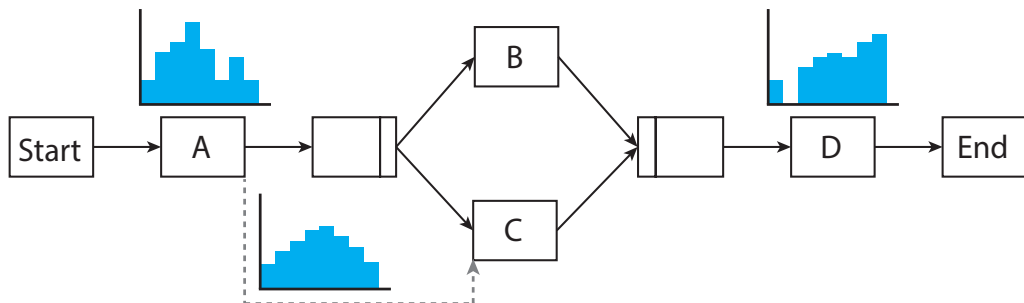


Abbildung 4.5: Prozess mit Histogramm Annotationen

Nach den zeitbasierten Histogrammen werden weitere grafische Darstellungs- und Analyse-Möglichkeit aus der Statistik in Verbindung mit zeitbasierten Daten vorgestellt.

4.3 Zeitbasierte Boxplots

Wie in Kapitel 2 beschrieben besitzt eine Aktivität eine Dauer und zwischen Aktivitäten existieren zeitliche Abstände. Die daraus resultierende zeitbehafteten Werte, lassen sich durch statistische Verfahren analysieren und beschreiben. Das statistische Verfahren der Boxplots, gibt eine visuelle Beschreibung der Messreihe von Zeitaspekte einer Aktivität. Durch diese Erstellung des Boxplots, lassen sich Informationen über die Messreihe der Daten bestimmen. Aus dem Boxplot kann das Minimum, Maximum, der genaue Median, das Lagemaß, die Ausreißer und die Streuung abgelesen werden. Daher eignet es sich sehr für die Analyse- und Darstellungs-Möglichkeit von zeitbehafteten Geschäftsprozessen. Dieses Verfahren der Analyse von zeitbasierten Boxplots findet auch im zweiten Schritt der Transformation des ETL-Prozesses statt (siehe Abbildung 4.1) und stellt die analysierten Daten für das Loading bereit.

4.3.1 Erstellung von zeitbasierten Boxplots

Analog zum Analyse-Verfahren der zeitbasierten Histogramme findet auch die Transformation der zeitbasierten Boxplots im zweiten Schritt der Transformation des ETL-Prozesses statt. Nach der selben Aufbereitung der Zeitstempel und der Verknüpfung der Informationen aus den Prozess-Event-Log können Dauern von Aktivitäten und Abstände zwischen Aktivitäten bestimmt und als Messreihe für die Analyse bereitgestellt werden. Diese Messreihen werden aufsteigend sortiert und die Quartile, der Interquartilsabstand und die schwachen Antennen bestimmt werden (siehe Kapitel 2). In den folgenden Abschnitten wird die Anwendung von zeitbasierten Boxplots auf die Dauer von Aktivitäten, Abstände zwischen Aktivitäten und selbst deklarierte Abstände erläutert.

Zeitbasierte Boxplots mit der Dauern von Aktivitäten

Die Dauer einer Aktivität ist definiert durch die Differenz des Endzeitpunktes (Zustandsänderung von *Started* in *Completed*) und des Anfangszeitpunktes (*Activated* in *Started*) einer Aktivität. Diese Differenz der Zustandsänderungen von vergangenen Aktivitäten können aus dem Event-Log berechnet und als Messreihe angesehen werden. Auf diesen Messreihen lassen sich Quartile, Median, Antennen über den Interquartilsabstand, wie in Kapitel 3 gezeigt, berechnen und bestimmen. Die Aussage der Streuung lässt sich durch die Verschiebung des Medians und den Boxen schließen. Ist die Box des Boxplots ungefähr mittig durch den Median geteilt, liegt höchstwahrscheinlich eine Normalverteilung vor und die Ausführung ist positiv zu bewerten. Ist der Median aber nach links oder rechts verschoben, lässt sich die Aktivität optimieren, da einige Ausführungen länger brauchen als die typische Ausführungszeit einer Aktivität. In Abbildung 4.6 sieht man ein optimales Boxplot und ein verschobenes Boxplot mit den Antennen, Ausreißern und ihren weiteren Eigenschaften. Die Ausreißer, die außerhalb der Antennen liegen, werden bei Boxplots direkt bei der Erstellung berücksichtigt und erkannt.

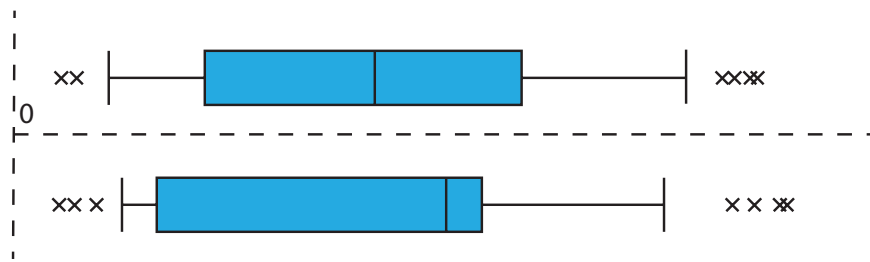


Abbildung 4.6: Optimales und verschobenes Boxplot

Zeitbasierte Boxplots mit Abstände zwischen Aktivitäten

Neben der Dauer von Aktivitäten lassen sich weitere zeitbasierte Differenzen zwischen Prozesselementen bestimmen. Der Abstand zwischen Anfangs- und Endpunkten zwischen Aktivitäten kann analog zu der Dauer gemessen und berechnet werden. Hierbei existieren zwischen zwei Aktivitäten vier Möglichkeiten den Abstand zu berechnen. Diese Abhängigkeiten wurden in Kapitel 2.2 vorgestellt und erläutert. In dieser Analyse-Möglichkeit wird nur die *END-START*-Abhängigkeiten zwischen Aktivitäten bewertet, da die drei weiteren Abhängigkeiten transitiv mit den Dauern der Aktivitäten und der *END-START*-Abhängigkeit verhalten. Dieser Abstand zwischen zwei Aktivitäten lässt sich analog durch ein Boxplot darstellen und analysieren. Der Wissensgewinn von Boxplots in Verbindung mit Dauern von Aktivitäten und den Boxplots der zeitlichen Abstände ist der selbe.

Neben den Abständen zwischen Aktivitäten ist der Abstand zum Prozessanfang für die Analyse- und Darstellungs-Möglichkeit interessant. Dieser Abstand lässt Aussagen über komplette Prozessabschnitte zu. Diese Abstände vom Prozessanfang können zum Aktivitäts-Anfang und Ende berechnet und analysiert werden. In dem *ADEPT2*-Modell werden diese Abstände zum Instanzanfang durch die Differenz des Startes oder Endes der Zielaktivität und dem Ende der Start-Aktivität des Geschäftsprozesses beschrieben. Zuletzt können noch Abstände zwischen Aktivität berechnet werden, die nicht direkt hintereinander liegen. Dieser Abstand wird nur berechnet, wenn dieser im Prozess durch eine Zeitkante separat deklariert wurde. Durch diese drei verschiedenen Abstände

wurden die wichtigsten Zeitaspekte abgedeckt und können für die Visualisierung der zeitbehafteten Daten verwendet werden.

Entscheidungen, Parallelität und Schleifen

In den wenigsten Fällen besteht ein Prozess aus sequenziell angeordneten Aktivitäten. Daher müssen Parallelitäten, Entscheidungen und Schleifen in der Analyse- und Darstellungs-Möglichkeit auch beachtet werden. Die Besonderheiten von Parallelitäten, Entscheidungen und Schleifen ist, dass diese durch eine Split- und eine Join-Aktivität umschlossen sind. Diese Split- und Join-Aktivitäten sind Aktivitäten, die vom System bearbeitet werden und daher eine Ausführungszeit von nahe null besitzen. Da die Dauer der nicht normalen Aktivitäten in den meisten Fällen immer gleich ist und daher nicht in das Gewicht fällt, werden diese bei der Berechnung von Abständen weggelassen und übergangen. Dadurch werden die Abstände von aufeinanderfolgenden Aktivitäten neu gesetzt. Falls der Nachfolger einer normalen Aktivität eine Split- oder Join-Aktivität ist, werden die Nachfolger der Split- oder Join-Aktivität für die normale Ausgangsaktivität gesetzt und der Abstand zwischen den neuen Nachfolgern berechnet. Die Abstände und Dauern in den Zweigen zwischen der Split- und Join-Aktivität wird unter den gleichen Bedingungen wie in einer Sequenz bestimmt. Die Analyse von Parallelitäten und Entscheidungen sind in diesem Modell die selben. Da der Abstand von zwei aufeinanderfolgende Aktivitäten bestimmt und die Join/Split-Aktivität eliminiert und überspannt werden. Dadurch fällt die Information bezüglich der Art des Blocks heraus. Bei Schleifen jedoch, wird zwischen den einzelnen Iterationen einer Aktivität unterschieden und nur die Daten analysiert, die die selbe Iteration besitzen. Die Ausnahme sind: Der Abstand der letzten Aktivität in einer Schleife zur ersten Aktivität in der nächsten Iteration und der Abstand zur Folgeaktivität nach der Schleife, welcher die höchste Iteration der Aktivität im Prozess-Log zur Hilfe nimmt. Auf diesen angepassten Reihen von zeitbehafteten Daten von Dauern und Abständen lassen sich die Boxplots wie in den vorigen zwei Abschnitten anwenden. Im nächsten Abschnitt wird die Annotation der Boxplots an die Prozesse gezeigt und Probleme bei der Analyse beschrieben.

4.3.2 Darstellung der Boxplots in Prozessen

Nach der Analyse der zeitbehafteten Daten können diese im letzten Schritt des ETL-Prozesses dem PMS zur Verfügung gestellt werden. Diese Analyse-Daten sind durch den Wert der Antennen, Werte der Box und den Median des berechneten zeitlichen Abstand oder Dauer definiert. Diese Informationen werden in dem analysierten Prozess an das jeweilige Prozesselement annotiert. Für eine Sequenz an Aktivitäten ist das in Abbildung 4.7 zu sehen. An die Aktivität C wurde das Boxplot annotiert und beschreibt ein nach rechtsverschobenes Boxplot. Zudem wurden zum Instanzstart die Abstände für den Benutzer visualisiert. Neben diesen einfachen Darstellungs-Möglichkeiten, wird in Abbildung 4.8 die komplexere Darstellungen mit Entscheidungen und Schleifen gezeigt. Bei der Darstellung der Entscheidung wird, wie in den obigen Abschnitten beschreiben, der Abstand vom Inneren der Entscheidung zur nächsten Aktivität nach der Join-Aktivität gezogen. Bei den Schleifen werden die verschiedenen Iterationen hinzugefügt und vor das Boxplot annotiert.

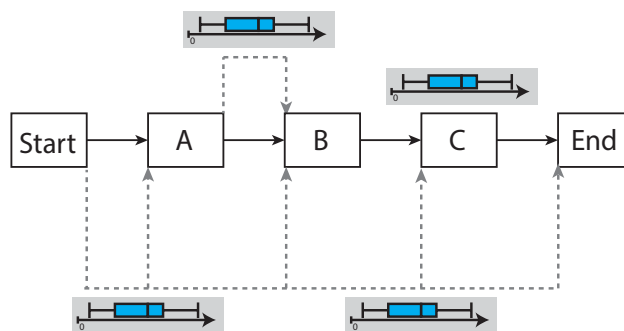


Abbildung 4.7: Prozess mit Boxplot Annotationen

Im folgenden Abschnitt werden Probleme dieser Analyse- und Darstellungs-Möglichkeit der Boxplots vorgestellt.

4.3.3 Probleme

Wie im vorigen Abschnitt angesprochen, gibt es bei diesem Verfahren keine Unterscheidung bei der Analyse von Folgeaktivität nach einer XOR- und Parallel-Join-Aktivität.

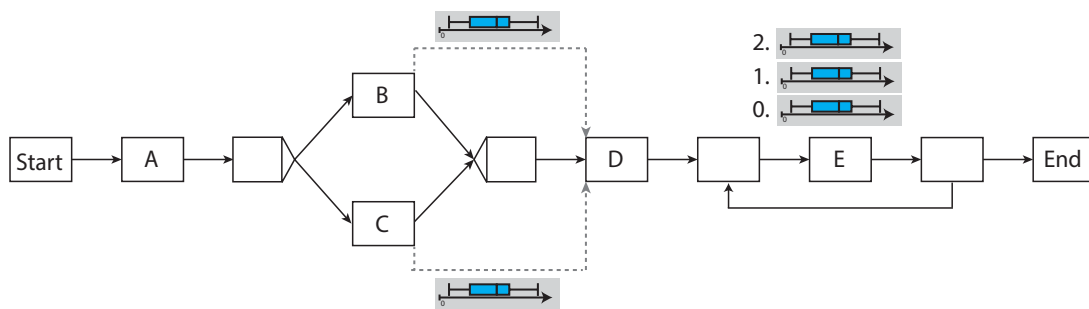


Abbildung 4.8: Prozess mit Entscheidungen und Schleifen

Daher ergibt sich bei verschieden lang dauernden Zweigen eines XOR-Block das Problem, dass die Folgeaktivitäten der Join-Aktivität mathematisch richtig analysiert werden, aber trotzdem falsche Analyse-Ergebnisse liefert. Wenn zusätzlich der kurze Pfad selten ausgeführt wird, werden diese Werte die außerhalb der Antennen liegen als Ausreißer deklariert. Zudem können in Entscheidungen bestimmte Eingaben im weiteren Prozess freigeschaltet werden. Durch diese Freischaltung von Elementen in einer Aktivität kann die Dauer einer Aktivität anwachsen und auch als Ausreißer deklariert werden, obwohl dies keiner ist. Diese beiden beschriebenen Beispiele sind in Abbildung 4.9 zu sehen und dargestellt. Der obere Zweig wurde wesentlich häufiger ausgeführt im Vergleich zu dem unteren. Somit wird die Ausführung der Aktivität über den unteren Zweig als Ausreißer analysiert und deklariert. Das selbe Problem entsteht bei Abständen, welche eine Schleife überspannen. Dieser Abstand wird je nach Iterationen-Anzahl um einiges größer und wird auch als Ausreißer erkannt und dargestellt.

Um diese Probleme der Entscheidungen und Schleifen zu beheben und Aktivitäten eindeutig zu identifizieren, müssen weitere Prozesseigenschaften aus dem PMS herangezogen und der Prozess transformiert werden. Dies wird im nächsten Abschnitt beschrieben.

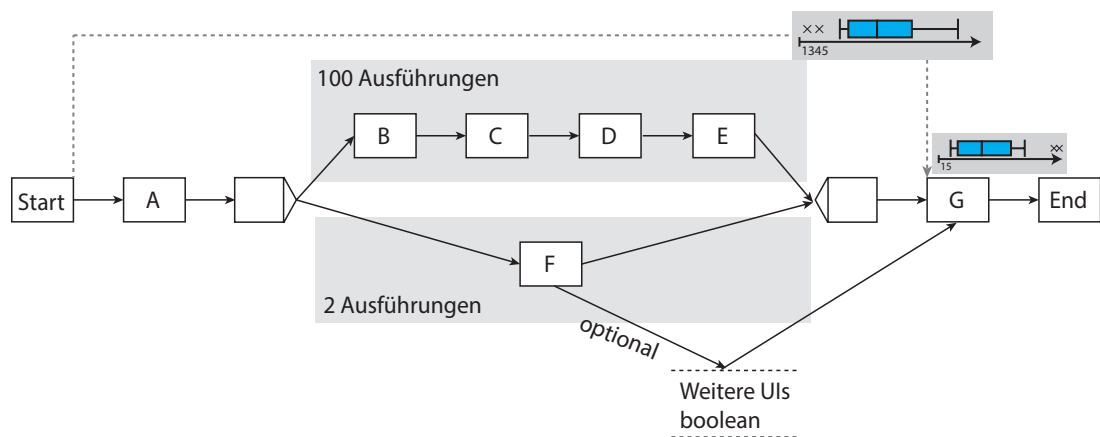


Abbildung 4.9: Problematische Abstände und Dauern

4.4 Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften

Prozesse laufen in den wenigsten Fällen sequenziell ab. Durch Entscheidungen und Schleifen existiert nicht nur ein Weg durch einen Prozess, sondern mehrere. Daher stellt sich die Frage, ob gleiche Aktivitäten mit unterschiedlichen ausgeführten Vorgängern überhaupt zusammen in einer Messreihe analysiert werden dürfen. Wie im vorigen Abschnitt beschrieben, müssen selten ausgeführte Zweige korrekt dargestellt und berechnet werden. Daher müssen gleiche Aktivitäten je nach ihrer Vorgeschichte in der Instanz separat behandelt und analysiert werden. Das selbe Problem existiert auch bei mehreren Schleifendurchläufen. Alle Start- und Endpunkte aller Folgeaktivitäten einer Schleife können je nach der Iterationenanzahl große Unterschiede aufweisen. Diese Folgeaktivitäten müssen je nach Iterationenanzahl auch separat analysiert werden. Um diese Eigenschaften aus den Anforderung des Kapitels 1 darzustellen, muss das Verfahren der Boxplots um die Prozesseigenschaft der Ausführungspfades erweitert und der Prozessgraph umgewandelt werden. Dies wird im Folgenden gezeigt.

4.4.1 Transformation des Prozesses und seiner Instanzen

Wie in Kapitel 1 beschrieben, soll die Analyse und Darstellung jede mögliche Kombination aus Sequenzen, Parallelitäten, Entscheidungen und Schleifen semantisch korrekt darstellen und analysieren. Da dies auf den vorhandenen *ADEPT2*-Prozessgraphen mit dem Prozess-Event-Log nicht einfach realisierbar ist, wird aus dem Prozessmodell ein Zeitmodell, wie in Abschnitt 3.5.1, erstellt. Diese Erstellung des Zeitmodells wird im zweiten Schritt der Transformation des ETL-Prozesses erledigt. Die Aktivitäten werden in zwei Knoten gesplittet, die Kontrollkanten werden von dem zweiten Knoten zum ersten Knoten der Folgeaktivität gezogen und die Dauer wird durch eine Kante der beiden gesplitteten Knoten dargestellt. Selbst deklarierte Zeitkanten werden von den Knoten der Ursprungsaktivität zur Zielaktivität angebracht und nicht normale Aktivitäten werden aus dem Graphen entfernt und die Nachfolger werden wie im vorigen Abschnitt angepasst. Diese Transformation ist in Abbildung 4.10 im ersten Schritt zu sehen.

Im nächsten Schritt wird für die genaue Erkennung der Ausführungspfade die Annotationen des eindeutigen Standorts im Zweig des Prozesses (engl. BranchID) an die Knoten hinzugefügt. Dadurch ist Definition des Knotens durch $Knoten(ID, BranchID, EventID, < (SuccID, EventID) >)$ im Zeitmodell gegeben. Die EventID stellt hierbei den ersten oder zweiten Knoten der gesplitteten Aktivität dar und kann die Zustände *start* oder *end* einnehmen. Weiter werden die Nachfolger des Knotens durch eine Aufzählung in der Definition des Knoten hinzugefügt, um den Abstand zwischen den Knoten später zu berechnen können. Diese Annotation an den Knoten ist in Abbildung 4.10 im Schritt 2 zu sehen. Es werden neben Aktivitäten auch Entscheidungen und Wiederholungen transformiert und annotiert.

Im letzten Schritt werden die Informationen aus dem Prozess-Event-Log auf das neue Zeitmodell angewandt. Hierbei wird für jede Aktivität bzw. Knoten einer Instanz der genaue Ausführungspfad bestimmt. Der Ausführungspfad (BranchIDs) wurde durch alle BranchID der vorigen ausgeführten Knoten bestimmt und dem eigenen Ausführungspfad mit der hinzugefügten eigenen BranchID des Knotens erweitert. In diesen BranchIDs sind keine Duplikate erlaubt, da die BranchID eines Pfades einmalig im Prozess ist und der Standort durch die vorigen Abfolgen von Aktivität eindeutig identifizierbar ist. In Abbil-

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

dung 4.10 im Schritt 4 werden die unterschiedlichen BranchIDs für den gleichen Prozess mit zwei verschiedenen Ausführungen und Instanzen gezeigt. Im oberen Beispiel wurde die BranchID 2 des obigen Pfad zu der folgenden BranchIDs hinzugefügt und im unteren die BranchID 3. Dadurch wird jede Aktivität trotz unterschiedlicher Ausführung eindeutig identifizierbar und korrekt analysierbar. Für den ersten Knoten der gesplitteten Aktivität mit der ID 7 wird in Abbildung 4.10 ein Beispiel mit den separierten Zeitpunkten aus dem Prozess-Event-Log und dem passenden Ausführungspfad der BranchIDs gezeigt. Das ist die weitere Sortierung und Separierung, welche vor der korrekten Analyse durchgeführt werden muss.

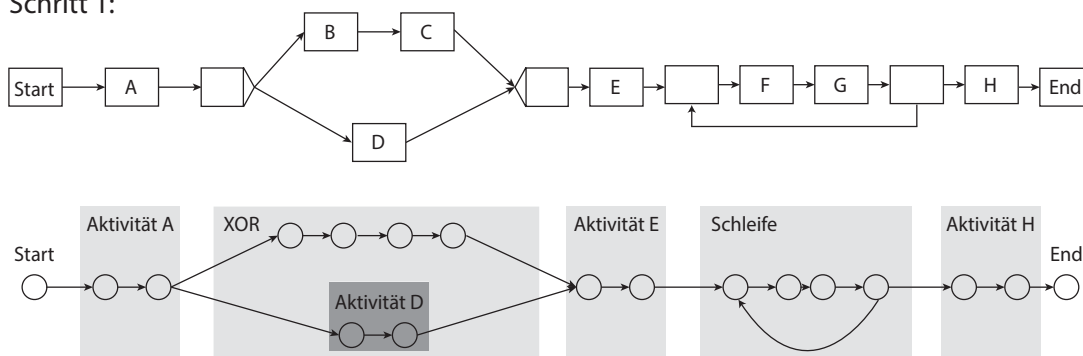
Im Folgenden wird aufbauend auf der Transformation und den Informationen der Ausführungspfade die Analyse der zeitbasierten Boxplots vorgestellt und erläutert.

4.4.2 Dauer und Abstände von Aktivitäten

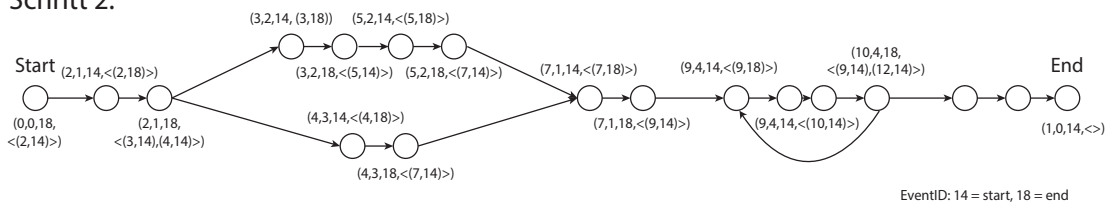
Nach der Transformation des Prozessgraphen in ein Zeitmodell muss man das vorhandene Zeitmodell wieder auf ein Prozessmodell zurück transformiert, um Abstände zwischen Aktivitäten und Dauern von Aktivitäten zu erhalten. Um die Dauer einer Aktivität in einem Prozess mit den passenden Ausführungspfad zu erhalten wird die Differenz der Zeitpunkte der beiden gesplitteten Knoten berechnet werden. Hierbei muss die InstanzID, die AktivitätID und die BranchIDs beider Knoten übereinstimmen. Sind diese Eigenschaften erfüllt, wird von dem Zeitpunkt des Knotens mit der EventID *end* der Zeitpunkt des Knotens mit der EventID *start* abgezogen. Durch Anwendung auf allen Instanzen eines Prozesses erhält man für alle vorhanden BranchIDs die passenden Dauern einer Aktivität zurück. Diese Dauer einer Aktivität wird durch $Akt(ActID, BranchIDs, EndTime - StartTime)$ definiert und kann durch diese Separierung für die Analyse weiter verwendet werden. Diese Separierung ist in Abbildung 4.11 zu sehen. Aus den unterschiedlichen Ausführungspfade und den beiden Knoten mit den verschiedenen EventIDs wird die Dauer zurück berechnet. Als Nebeneffekt dieser Separierung kann die relative und absolute Häufigkeit der unterschiedlichen Ausführungspfade bestimmt und zusätzlich an die Boxplots hinzugefügt werden. Diese Informationen sieht man in Abbildung 4.11.

4.4 Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften

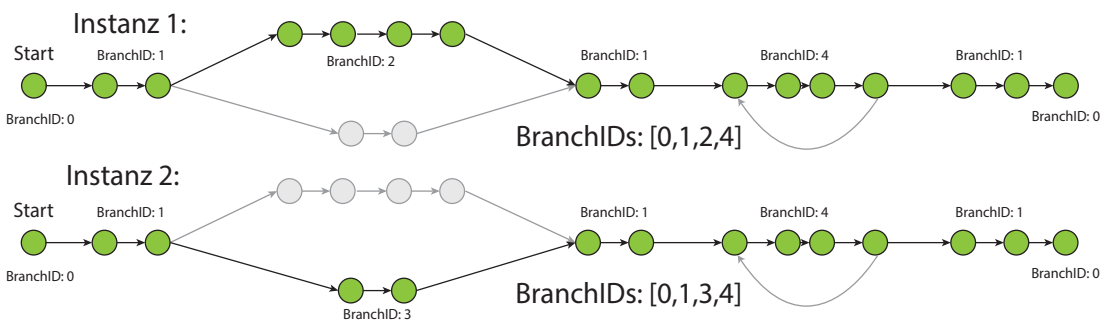
Schritt 1:



Schritt 2:



Schritt 3:



Beispiel: Knoten(7,0,14):

1. BranchIDs [0,1,2,4]: 1234,2304,5098
2. BranchIDs [0,1,3,4]: 501,899,495

Abbildung 4.10: Transformation des Prozessgraphen und der Instanzen

Bei Abständen zwischen Aktivitäten muss beachtet werden was die Quelle und das Ziel des Abstandes ist. Dieser Abstand ist definiert durch $Edge(QuellID, QuellEvent, ZielID, ZielEvent, ZielBranchIDs, ZielTime - QuellTime)$. Dieser Abstand kann nur berechnet werden, wenn die beiden Knoten aus der selben Instanz stammen und die BranchIDs

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

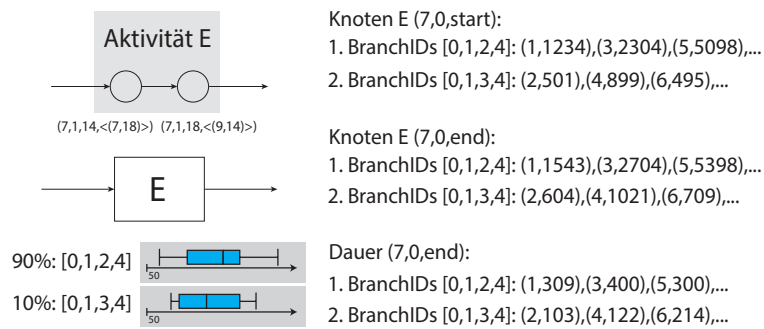


Abbildung 4.11: Berechnung der Dauer einer Aktivität

der Quelle alle in der BranchIDs des Ziels vorkommen. Dieses Vorgehen ist in Abbildung 4.12 zu sehen und die Berechnung der Differenz wurde anhand des Abstandes von dem Ende von Aktivität A zum Anfang von der Aktivität E berechnet. Die Erstellung des Boxplots läuft analog zu der Dauer ab und wird durch die Häufigkeit ergänzt.

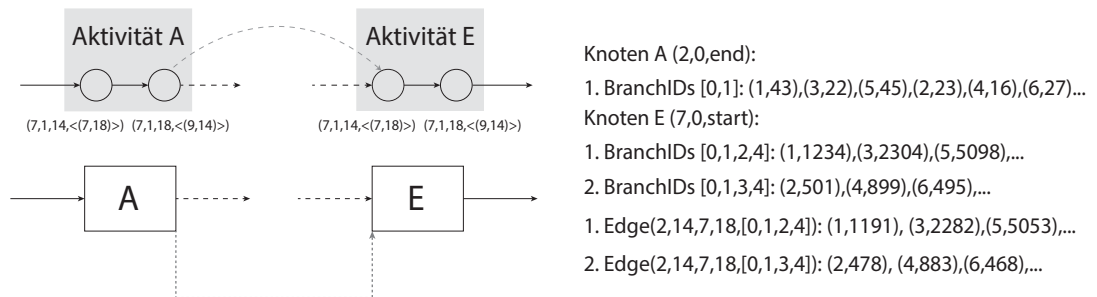


Abbildung 4.12: Berechnung des Abstands zwischen Aktivitäten

Durch diese Anpassungen sind die meisten Anforderungen aus dem Kapitel 1 erfüllt. Im nächsten Abschnitt werden zuletzt die Wiederholungen von Prozesselementen erläutert.

4.4.3 Schleifen in Graphen

Zuletzt müssen Wiederholungen von Prozesselementen abgebildet und korrekt berechnet werden. Hierbei kann die Dauer der selben Aktivität in einer weiteren Iteration von der

4.4 Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften

vorigen Iteration essenziell abweichen. Daher müssen sich wiederholende Aktivitäten im Ausführungspfad von den vorigen Ausführungen unterscheiden. Zudem muss diese Unterscheidung für spätere, nach der Schleife folgende Aktivitäten übernommen werden. Um diese Separierung von mehreren Iterationen im Ausführungspfad darzustellen, müssen die BranchIDs aus dem vorigen Abschnitt durch Hochzahlen an der betroffenen BranchID der Wiederholung, erweitert werden. Bei einer Wiederholung wird dabei, wie in Abbildung 4.13 zu sehen, die BranchID der Wiederholung um eins erhöht. Die neuen BranchIDs werden für die Ausführungspfade in der Schleife und für alle folgenden Aktivitäten außerhalb der Schleife eindeutig identifizierbar. Dadurch sind verschiedene Wiederholungen von Prozesselementen semantisch korrekt analysierbar und können später in der Darstellung visualisiert werden.

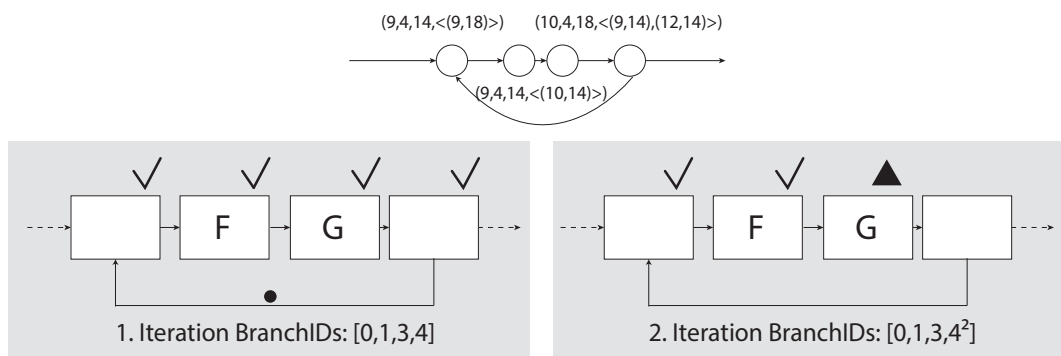


Abbildung 4.13: Schleifen und in Verbindung mit den BranchIDs

4.4.4 Zusammenfügen von verschiedenen BranchIDs

Durch das obige Verfahren, werden wie in den vorigen Abschnitten beschrieben alle vorhandenen Möglichkeiten, welche im Prozess-Event-Log vorkommen, analysiert und für die Darstellung bereitgestellt. Solch ein Prozess mit allen Möglichkeiten ist in Abbildung 4.14 zu sehen. Jedoch kann ein Prozess mit vielen Entscheidungen und Wiederholungen einige Möglichkeiten kreieren. Um die Übersichtlichkeit für den Benutzer zu erhalten, wird versucht ungefähr gleiche Boxplots eines selben Prozesselementes mit unterschied-

4 Analyse und Darstellung von zeitbehafteten Geschäftsprozessen

licher BranchIDs in ein einziges Boxplot zusammenzuführen. Jedoch darf bei allen folgenden Prozesselementen des jeweiligen Ausführungspfad eine geringe Abweichung zwischen allen Boxplots entstehen. Dies darf bei jeder vergleichbaren Boxplot-Grenze eines Prozesselementes (Antennen, Median, Quartile) nicht mehr als 5% betragen. Ist die Abweichung gleich oder geringer, wird wie in Abbildung 4.14 gezeigt, die Pfade zusammengefügt, die Schnittmenge beider BranchIDs gebildet und für das Element gesetzt. Dadurch wird die Übersichtlichkeit für den Benutzer verbessert und die Anzahl der anzeigenden Möglichkeiten im Prozess verringert.

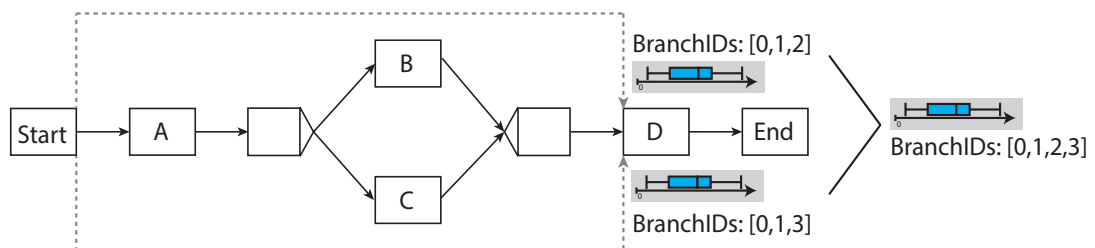


Abbildung 4.14: Darstellung der Zusammenführung von Boxplots

Neben den Analysen der Prozesselemente lassen sich durch die Betrachtung von den Start- und Endzeitpunkte der Aktivitäten Aussagen bezüglich der Organisationsstruktur gewinnen. Diese Gewinnung der bevorzugten Ausführungszeiträume von Aktivitäten wird im nächsten Kapitel beschrieben.

4.5 Analyse des Ausführungszeitraumes einer Aktivität

Für die Vorhersage einer korrekten Ausführung einer Instanz können aus den historischen Daten die Zeiträume von bevorzugten Ausführungen einer Aktivität analysiert werden. Um diese Zeiträume zu erhalten werden alle Aktivitäten mit der gleichen Definition der auszuführenden Benutzer (engl. Staff Assignment Rule [RCWR04]) an einer Zeitlinie annotiert. Diese Zeitlinie ist wie ein Histogramm durch Intervalle unterteilt. Diese Größe des Intervalls wird durch die Hälfte des kleinsten Medians aus den Verfahren der Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften aller

4.5 Analyse des Ausführungszeitraumes einer Aktivität

Aktivitäten mit der gleichen Staff Assignment Rule gewählt. Nach der Bestimmung des Intervalls, werden die Dauern die diese Intervalle überspannen gezählt. Diese Annotation der Dauern mit den Start- und Endzeitpunkten ist in Abbildung 4.15 für die Aktivität A zu sehen. Nach dieser Annotation wurden die Aktivitäten, welche in diesem Zeitraum stattfinden gezählt. Nach dieser Zählung wird eine Bewertungsgrundlage für die Einteilung erstellt. Der größte Wert der Zählung wird ermittelt und dieser als *grün/ausführbaren* Bereich definiert. Analog wird von diesem maximalen Wert ausgehend die Intervalle welche 25% des Maximum beinhalten als *gelb/kritisch* und Intervalle die weniger 5% des Maximum als *rot/nicht ausführbar* deklariert. Diese Einteilung und Bewertung ist für die Aktivität A und B, welche beide aus unterschiedlichen Benutzergruppen bestehen, in Abbildung 4.15 zu sehen und dargestellt.

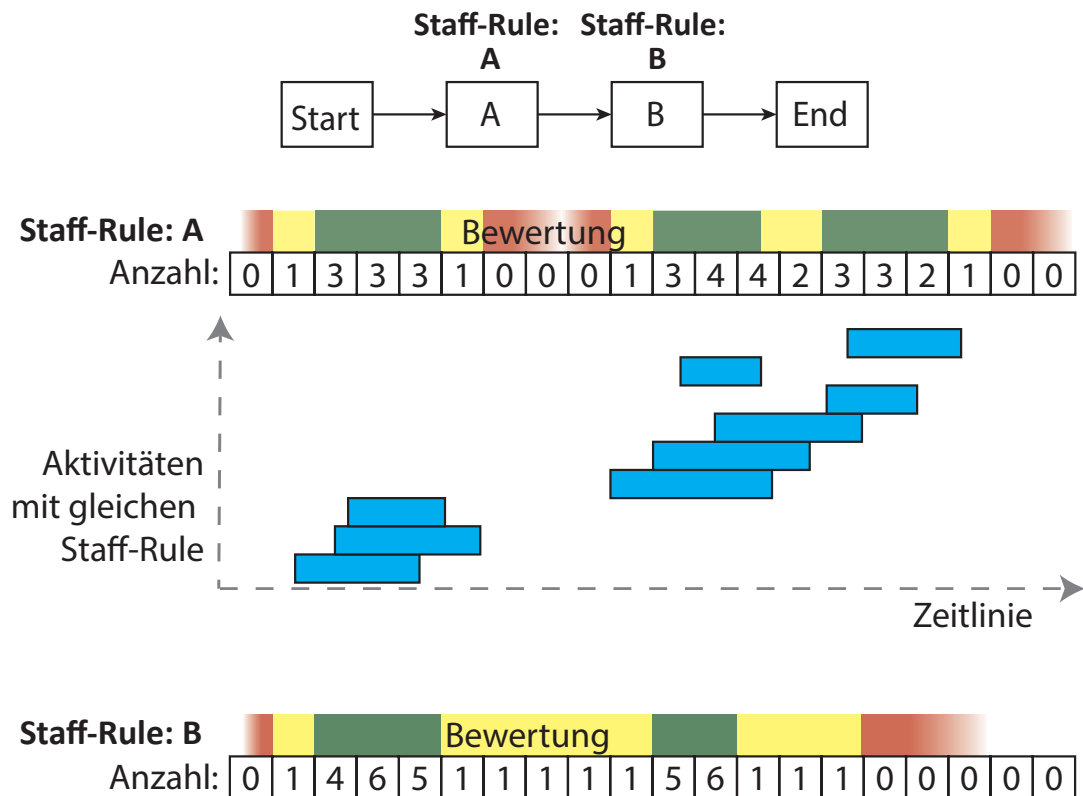


Abbildung 4.15: Analyse des Ausführungszeitraumes

Im letzten Abschnitt wird ein Vergleich und Fazit der vorgestellten Verfahren gezogen und diese anhand den Anforderungen aus Kapitel 1 bewertet.

4.6 Vergleich und Fazit

Nach der Vorstellung von einzelnen Analyse- und Darstellungs-Möglichkeiten von zeitbehafteten Geschäftsprozessen kann jedes Verfahren anhand der Eigenschaften, die sie erfüllen, bewertet werden. Hierbei müssen alle drei Verfahren die Anforderungen aus dem Kapitel 1 erfüllen. Die Analyse und Darstellung der Dauer von Aktivitäten, Abstände zwischen Aktivitäten und Abstände zum Instanz-Anfang wird von allen drei vorgestellten Verfahren unterstützt. Jedoch muss man bei den Verfahren der Histogramme und den Boxplots ohne Transformation Einschnitte bezüglich der semantische Korrektheit machen. Selten auftretende Ausführungen in Entscheidungen, Schleifen und optionale Dinge in Prozesselementen wurde in den Verfahren der zeitbasierten Histogrammen und Boxplots richtig berechnet, aber waren Semantisch nicht richtig. Deswegen musste das Prozessmodell in ein Zeitmodell umgewandelt und mit weiteren Prozesseigenschaften aus dem PMS angereichert werden. Nach dieser Umwandlung konnten alle Ausführungspfade separiert und die Häufigkeit der Ausführungen bestimmt werden. Durch diese Separierung konnte man die semantische Korrektheit von Wiederholungen, optionale Elemente und Entscheidungen erreichen. Daher wurden alle Anforderungen aus Kapitel 1 im letzten Verfahren der Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften umgesetzt. Auf diesen angepassten Messreihen kann das Verfahren der zeitbasierten Histogramme auch angewandt werden, jedoch muss für Bewertungsgrundlage und Korrektheit noch größerer Aufwand betrieben werden. Daher wurden die Boxplots als Schwerpunkt für das Monitoring verwendet und für die Darstellung eine Lösung präsentiert, um die Übersichtlichkeit der Darstellung für zeitbasierten Geschäftsprozessen zu gewährleisten. Zuletzt wurde durch eine Abwandlung von Histogrammen die Ausführungszeiträume anhand der Organisationsstruktur von Aktivitäten vorgestellt. Dieses und das Verfahren aus Abschnitt 4.4 werden für das Monitoring im nächsten Kapitel verwendet um eine Echtzeitüberwachung der Instanzen und Aktivitäten zu ermöglichen.

5

Monitoring und Vorhersagbarkeit von zeitbehafteten Geschäftsprozessen

Die Überwachung und die Vorhersagbarkeit von den zeitbehafteten Geschäftsprozessen beschreibt die Ausführbarkeit und Korrektheit von Prozesselementen der Vergangenheit und der Zukunft. Diese Überwachung stellt sicher, dass eine Instanz und ihre Elemente zeitnah erledigt und durch ihre zeitbehafteten Ereignisse bewertet werden. Diese Überwachung von Instanzen und den Prozesselementen wird im folgenden Kapitel beschrieben. Im ersten Abschnitt des Kapitels wird das Zusammenspiel des PMS, der Analyse-Komponente und des Monitors vorgestellt. Die Monitoring-Komponente verwendet die Analyse-Möglichkeit der Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften für die Erstellung der Bewertungen einer Instanz und deren Elementen. Durch dieses verwendete Verfahren lassen sich alle Anforderungen aus Kapitel 1 für den Monitor erfüllen. Nach der Beschreibung der Zusammenhänge

der Komponenten wird jede einzelne Komponente näher beschrieben und definiert. Im letzten Abschnitt wird die Vorhersage der Bewertungen von Instanzelementen besprochen und das rückwirkende Monitoring für die Optimierung und Fehlerkorrektur im *PMS* beschrieben.

5.1 Zusammenspiel der Komponenten

Bei der Ausführung von Prozessen in einem Prozess-Managementsystems, werden Ereignisse (engl. Events) über die Zustandsänderungen von Prozesselementen generiert und in den Prozess-Event-Log geschrieben. Dieser Prozess-Event-Log beherbergt alle Informationen der Ausführung der Prozessinstanzen aus der Vergangenheit und wird für die Analyse der zeitbehafteten Geschäftsprozesse verwendet. Zudem teilt dieser Prozess-Event-Log dem Monitor die neuen Ereignisse aus den aktuellen Instanzen mit. Um diese Instanzen und die Elemente zu bewerten, stellt die Analyse-Komponente dem Monitor diese Analyse-Daten, die aus dem Analyse-Verfahren resultieren, zur Verfügung. Durch diese Analyse-Daten bewertet der Monitor die Ereignisse der Instanz bezüglich der Zeitaspekte. Die einzelnen Komponenten sind, wie in Abbildung 5.1 zu sehen, in weitere Einzelteile aufgeteilt. Dabei gibt es eine offline Analyse, die die zeitbehafteten Daten aus dem Prozess-Event-Log mit dem Verfahren aus Kapitel 4.4 beschreibt und eine online Analyse, die die Analyse-Daten des offline Monitoring anpasst und diese für die Vorhersagbarkeit von Prozesselementen nutzt. Weiter wird der Monitor selbst in ein online und offline Monitoring unterteilt. Diese unterschiedlichen Monitoring-Verfahren sind zum einen für die Echtzeitüberwachung zuständig (online) und zum anderen für die Bewertung von der Instanzen beim Auftreten eines Ereignisses (offline).

Im folgenden Abschnitt werden die unterschiedlichen Verfahren und die Bewertungen der Monitoring-Komponenten vorgestellt. Diese Monitoring-Komponente basiert auf den Analyse-Daten der Analyse aus dem vorigen Kapitel und erhält Informationen zu den laufenden Instanz aus dem Prozess-Event-Log eines *PMS*.

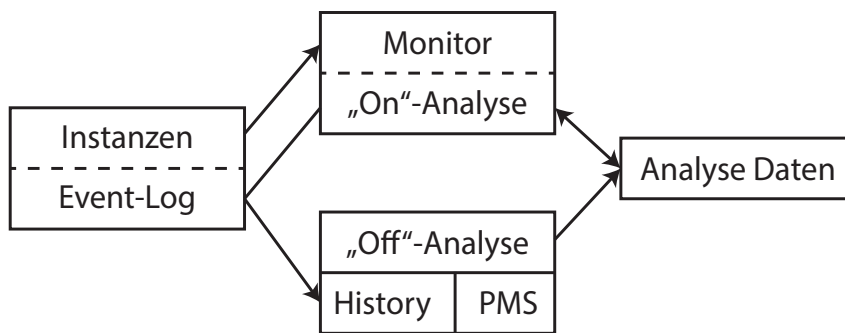


Abbildung 5.1: Zusammenspiel von Analyse, Daten-Logs und des Monitors

5.2 Online und Offline Monitoring

Wie im obigen Abschnitt beschrieben, ist der Monitor in zwei Bereiche unterteilt. Zum einen das offline Monitoring und zum anderen das online Monitoring. Beide Bereiche verwenden die Ereignisse aus dem Prozess-Event-Log und bewerten diese anhand der zeitbehafteten Analyse-Daten aus Kapitel 4.4. Beide Monitoring-Bereiche sind für separate Überwachungen von Prozessen zuständig. Das online Monitoring überwacht, wie in Abbildung 5.2 zu sehen die Dauer einer Aktivität, den Abstand zwischen aufeinanderfolgenden Aktivitäten und die deklarierten Abstände durch eine Zeitkante im Prozessgraphen. Diese Überwachung läuft in Echtzeit ab und bewertet die zeitlichen Attribute von Prozesselementen, die zu lange dauern. Dadurch wartet der Monitor auf das passende Ereignis eines Elementes aus der Instanz. Tritt dieses Element nicht rechtzeitig auf, wird das Element in Echtzeit bewertet. Das offline Monitoring überwacht, wie in Abbildung 5.3 zu sehen den Fortschritt einer Instanz und zu kurz dauernde Aktivitäten und Abstände. Diese Überwachung findet nicht in Echtzeit statt, da die Anzahl an Ereignissen sonst extrem zunimmt und die kritischen Momente durch die Verbindung von offline und online Monitoring abgebildet werden können. Diese kritischen Bewertungen sind durch den bewerteten Abstand des offline Monitorings zum letzten Ereignis der Instanz und der Bewertung des folgenden Abstands bzw. der Dauer aus dem online Monitoring abgedeckt. Deswegen wird diese Bewertung des Fortschrittes erst beim

5 Monitoring und Vorhersagbarkeit von zeitbehafteten Geschäftsprozessen

Auftreten des Ereignisses durchgeführt. Das offline Monitoring wartet im Vergleich zum online Monitoring also nicht auf die Ereignisse, sondern bewertet diese beim Auftreten.

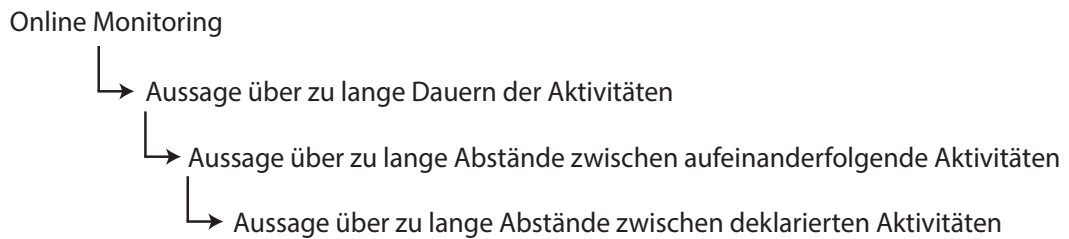


Abbildung 5.2: Aufbau des Online Monitorings

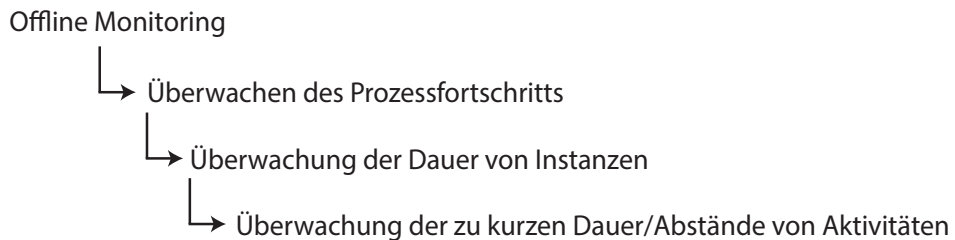


Abbildung 5.3: Aufbau des Offline Monitorings

Im folgenden Abschnitt wird die Bewertungsgrundlage bezüglich der Boxplots aus dem vorigen Kapitel und das Mapping der dazugehörigen Zustände erledigt.

5.2.1 Bewertungsgrundlage von Prozesselementen

Damit die Monitoring-Komponente eine automatisierte Bewertung von Prozesselementen durchführen kann, müssen Informationen aus der Vergangenheit analysiert und dem Monitor bereitgestellt werden. Bei diesen Bewertungen stellt sich die Frage, wie Analogien aus unserer Umwelt und der Gesellschaft für den Benutzer selbsterklärend dargestellt und bewertet werden können. Es stellte sich bei den Boxplots heraus, dass diese mit ihren mathematischen Grenzen eine Einteilung für den Benutzer und dem System ermöglichen. Es wurde festgestellt, dass diese Grenzen des Boxplots in drei

Bereiche eingeteilt werden kann, die für den Benutzer interessant sind und eine Bewertung der Ausführungszeit zulassen. Zum einen den komfortablen Bereich zwischen dem unteren und oberen Quartil, den kritischen Bereich zwischen den Quartilen und den schwachen Antennen und zuletzt die Bereiche der Ausreißer. Diese drei Bereiche kann man den Farben einer Ampel zuordnen. Die Farben hierfür sind *grün* für Zustand *korrekt*, *gelb* für *kritisch* und *rot* für *fehlerhaft*. Diese Zuordnung der Zustände zu den Bereichen des Boxplots ist in Abbildung 5.4 zu sehen und visuell aufbereitet.

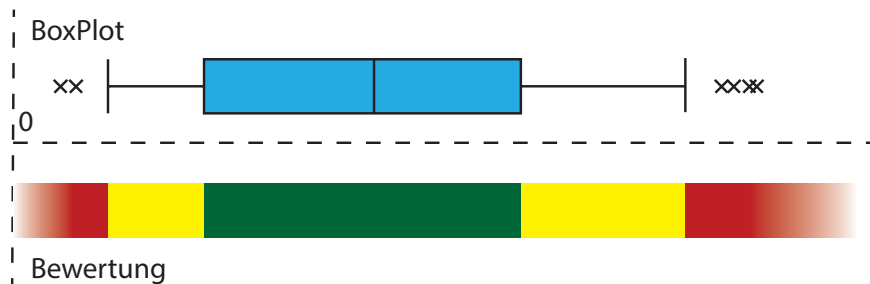


Abbildung 5.4: Boxplot mit Bewertung

Um nun Dauern von Aktivitäten und Abstände zwischen Aktivitäten in einem Monitor automatisiert zu bewerten und darzustellen, müssen Definitionen für die Zustände erstellt werden. Hierbei existieren für ein Prozesselement folgende Fälle bezüglich des zeitlichen Attributes:

- Dauer/Abstand ist zu groß,
- Dauer/Abstand ist wie erwartet,
- Dauer/Abstand ist zu klein,
- Dauer/Abstand ist im unteren Grenzbereich,
- Dauer/Abstand ist im oberen Grenzbereich.

Diese fünf Fälle müssen auf die Bereiche des Boxplot abgebildet werden. Diese mathematische Abbildung von Zeiten auf Zustände ist folgende:

- rot-fehlerhaft:

5 Monitoring und Vorhersagbarkeit von zeitbehafteten Geschäftsprozessen

- Dauer/Abstand ist zu groß
 - * Werte größer der oberen Antenne
- Dauer/Abstand ist zu klein
 - * Werte kleiner der unteren Antenne
- gelb-kritisch:
 - Dauer/Abstand ist im unteren Grenzbereich
 - * Werte zwischen dem unteren Quartil und der unteren Antenne
 - Dauer/Abstand ist im oberen Grenzbereich
 - * Werte zwischen dem oberen Quartil und der oberen Antenne
- grün-korrekt
 - Dauer/Abstand ist wie erwartet
 - * die Werte liegen zwischen dem unteren ($Q_{0,25}$) und oberen Quartil ($Q_{0,75}$)

Durch diese Einteilung und Zuordnungen lassen sich nun Bewertungen für die sequenzielle Prozesselemente erstellen. Bei Entscheidungen und Wiederholungen müssen mehrere Möglichkeiten des Ausführungspfades für das online und offline Monitoring beachtet werden. Bei diesen Prozesselementen spielt wieder der Ausführungspfad der BranchIDs aus den Analyse-Daten des Abschnittes 4.4 eine Rolle. Dazu wird für die selbe Aktivität mehrere Bewertungsbereiche definiert und je nach Ausführung anhand der BranchIDs bewertet. Ein bewerteter Prozess mit einer Entscheidung ist in Abbildung 5.5 zu sehen. In dem Diagramm für die Bewertung werden verschiedene Ausführungsmöglichkeiten einer Instanz beschrieben. Schleifen, die in der Analyse-Komponente bewertet und erkannt wurden, werden analog durch die unterschiedlichen BranchIDs identifiziert.

Nach den Bewertungsgrundlagen für die Prozesselemente wird die Monitoring-Komponente vorgestellt. Dieses offline und online Monitoring wird durch eine visuelle Darstellung für bewertete Prozesselemente dargestellt. Diese beiden Monitoring-Verfahren werden im Folgenden beschrieben. Die beiden Verfahren benutzen die Bewertungsgrundlage aus diesem Abschnitt für die Bewertung der weiteren Prozesselemente.

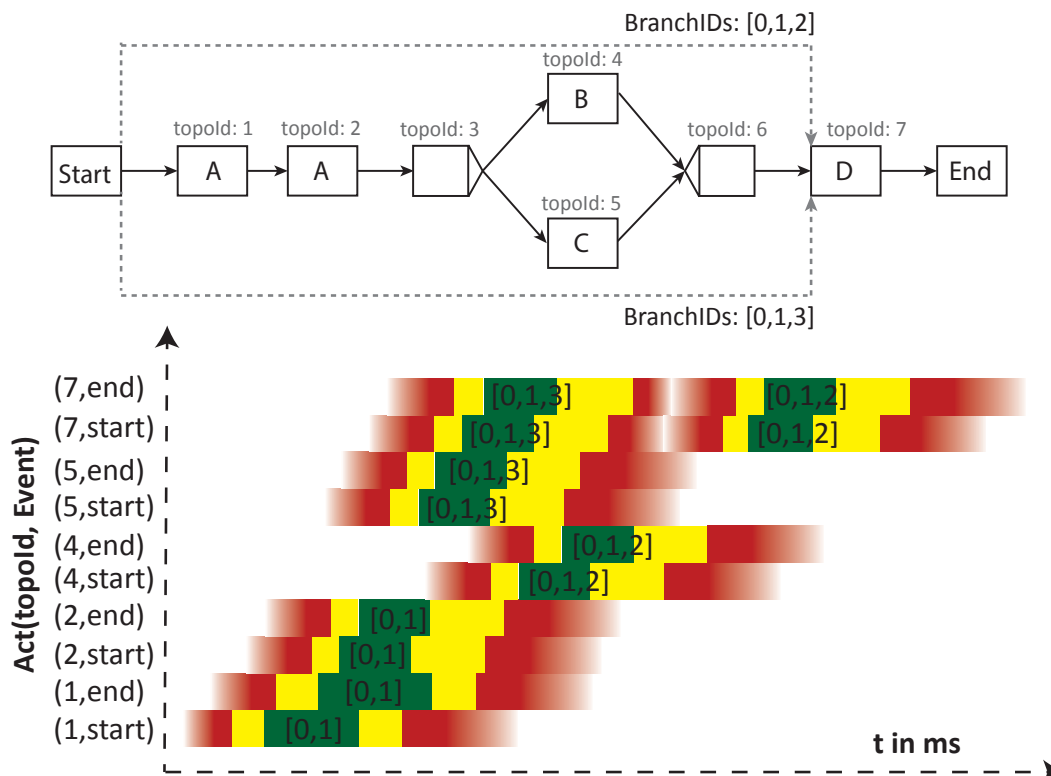


Abbildung 5.5: Monitoring von Entscheidungen

5.2.2 Online-Monitoring

Nach der Abbildung der Zustände können diese Bewertungen auf eine laufende Instanz angewandt werden. Dieses online Monitoring bewertet die Dauer von Aktivitäten und den Abstände von aufeinanderfolgenden Aktivitäten, welche zu lang dauern. In Abbildung 5.6 ist eine laufende Instanz mit fünf Aktivitäten, die zum Teil parallel ablaufen dargestellt. Die Aktivität A wurde in der vorgesehenen Zeit erfolgreich beendet (annotierter Haken an der Aktivität) und grün eingefärbt. Bei den Abständen zwischen der Aktivität A und den darauf folgenden Aktivitäten trat bei zwei Abständen Probleme auf. Der Abstand zwischen Aktivität A und Aktivität C wurde als kritisch eingestuft und der Abstand zur Aktivität D als fehlerhaft. Diese Abstände sind durch den Start der anschließenden Aktivität abgeschlossen (siehe gefüllter Kreis). Die Aktivitäten B,C und D sind gestartet

5 Monitoring und Vorhersagbarkeit von zeitbehafteten Geschäftsprozessen

und laufen im Moment (siehe Annotation des Dreiecks). Diese Aktivitäten werden durch das online Monitoring in Echtzeit überwacht und bewertet. Hierbei läuft die Aktivität B im fehlerhaften Zustand, die Aktivität D im kritischen. Diese beide Aktivitäten haben eine zu lange Dauer. Die Aktivität C liegt im vorgesehenen Bereich. Der endgültige Zustand wird jedoch erst beim Eintreten des Beenden der Aktivität festgestellt, daher kann die sie einen kritischen oder fehlerhaften Zustand beim Überschreiten der oberen Quartil-Grenze erreichen.

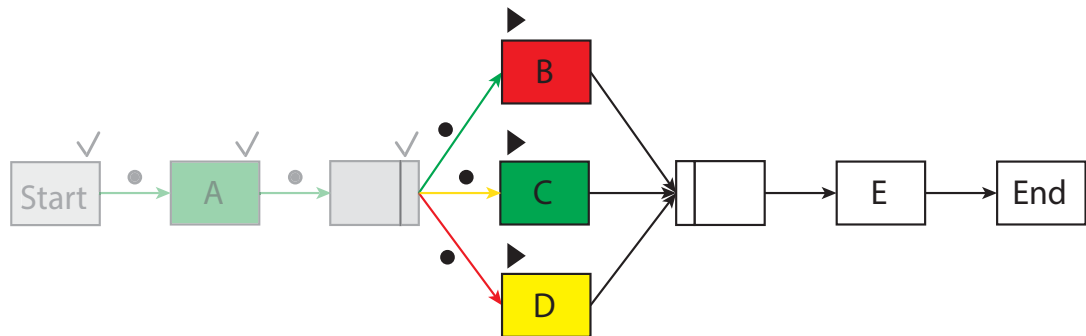


Abbildung 5.6: Laufende Instanz mit Bewertungen

5.2.3 Offline-Monitoring

Nachdem die zu große Dauern von Aktivitäten, Abstände von Folgeaktivitäten und deklarierte Zeitabstände durch das online Monitoring in Echtzeit überwacht werden, fehlt für die restlichen Eigenschaften der Fortschritt einer Instanz und die Behandlung von zu kurzen Dauern von Prozesselementen. Die Bewertung der Elemente läuft auch beim offline Monitoring durch die Bewertungsgrundlage aus Abschnitt 5.2.1 ab. Bei der Bewertung von dem Fortschritt einer Instanz ist zu beachten, dass eine Instanz mehrere Prozesselemente beinhaltet und dieser Fortschritt mehrere Elemente umfasst. Dieser Fortschritt wird durch den Abstand von dem Beginn der Instanz bis zum momentanen Ereignis des Prozesselements berechnet. Dabei besitzt eine Aktivität mindestens zwei Ereignisse und zwar das Start- und End-Ereignis. Der Zeitpunkt des Auftretens des Ereignisses kann für das offline Monitoring verwendet werden und mit den Analyse-

Daten, welche den Abstand zum Beginn der Instanz beinhaltet, verglichen werden. Hierbei kann die Instanz wieder mit den Zuständen *rot* für *fehlerhaft*, *gelb* für *kritisch* und *grün* für *korrekt* beschrieben werden. Da dies aber eine zu geringe Granularität vorweist und eine Instanz aus vielen Elementen bestehen kann, wird jeder Zustand in fünf weitere Teile gegliedert. Durch diese Unterteilung lässt sich eine Art Heatmeter erzeugen, welcher die Instanz bewertet. Dieser Heatmeter ist in Abbildung 5.7 zu sehen und der momentane Zustand wird durch einen Zeiger dargestellt. Dieser Heatmeter kann Werte von 0% links (grün) bis 100% fehlerhaft rechts (rot) einnehmen. Dazwischen liegt die kritische Phase (gelb) der Instanz. Um diesen Zustand zu berechnen, werden die Zustände der Abstände vom Beginn der Instanz addiert und durch die bisherigen Anzahl der aufgetretenen Ereignisse dividiert. Hierbei hat jeder Zustand einen eigenen Wert:

- Korrekt: 0 (grün)
- Kritisch: 1 (gelb)
- Fehlerhaft: 2 (rot)

Die Summe aller Zustände dividiert durch die Anzahl der aufgetretenen Ereignissen ergibt den Wert des Heatmeters. In Abbildung 5.7 ist das online und offline Monitoring in einem Diagramm dargestellt. Je nach Ausführung kann der Heatmeter in beide Richtungen ausschlagen. Daher können kritische oder auch fehlerhafte Instanzen nach einigen Prozesselementen in den grünen Bereich zurück gelangen.

Im Weiteren bewertet das offline Monitoring die zu kurze Dauer von Aktivitäten und zu geringe Abstände zwischen Aktivitäten. Dies kann im online Monitoring nicht semantisch Korrekt dargestellt werden, da es Probleme und Verwechslungen mit zu lang dauernden Prozesselementen gibt. Die Farbe rot ist hierbei zu kurz und zu lang dauernden Prozesselementen zugeordnet und eine eindeutige Zuordnung während der Ausführung ist nicht zu erkennen. Die Lösung dafür ist, eine eindeutige Zuordnung durch die weitere Einführung zweier Farben für die zu kurz dauernden Elementen. Somit könnten solche Aktivitäten auch im online Monitoring bewertet werden. Durch die Einteilung in zwei Bereiche des Monitorings lassen sich die zeitlichen Abhängigkeiten von Prozesselementen automatisiert überwachen. Anhand dieser Bewertung können anschließend weitere Aus-

sagen über die zukünftigen Zustände der Aktivitäten und Abstände vorhersagt werden. Dieses Verfahren für die Vorhersage, die online Analyse und das daraus resultierende rückwirkende Monitoring wird in den nächsten beiden Abschnitten vorgestellt.

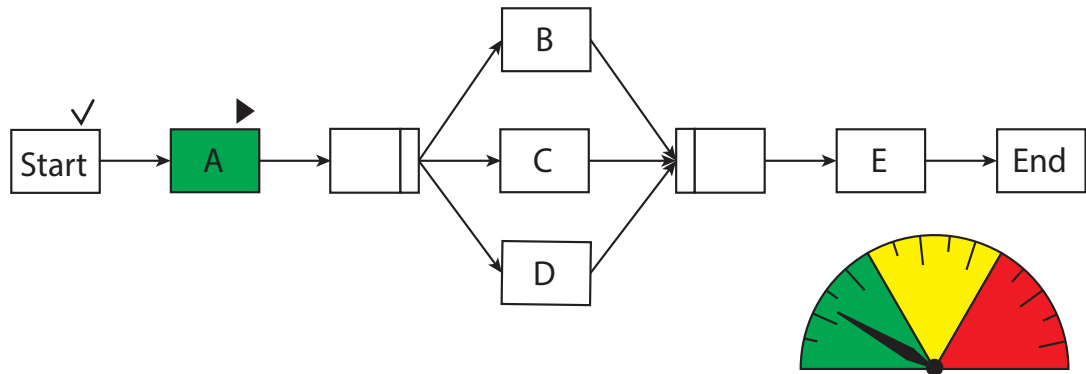


Abbildung 5.7: Online und Offline Monitoring

5.3 Vorhersage und online Analyse

Wie in Abschnitt 5.2 beschrieben, existiert neben der offline Analyse die online Analyse. Diese online Analyse passt die Analyse-Daten der offline Analyse während der Laufzeit an die Umgebung und der Organisationsstruktur im Unternehmen an. Durch diese Anpassung an das Umfeld werden die Grenzen der zeitbasierten Boxplots in Verbindung mit dem Verfahren aus Kapitel 4.5, der Analyse der Ausführungszeiträume angepasst. Hierbei werden Schnittmengen der Zustände des Boxplots und des Ausführungszeitraums gebildet. Danach findet durch die folgende Zuordnung der Zustand-Schnittmengen die Anpassung des Folgezustandes des Boxplots statt (dargestellt durch Zustand-Boxplot/Zustand-Ausführung): Für die Zustände *korrekt/korrekt* resultiert der Zustand *korrekt* im Boxplot, aus *korrekt/kritisch* resultiert *kritisch* und alle weiteren Kombinationen werden durch den Zustand *fehlerhaft* deklariert. Diese Anpassung und online Analyse findet beim Starten einer Instanz statt. Durch dieses Verfahren lassen sich bestimmte Zustände von Prozesselementen schon im Voraus bestimmen und vorhergesagen. Die

Schnittmenge der Zustände von Boxplots eines Prozesses und der Ausführungszeiträume der Organisationsstruktur ist in Abbildung 5.8 zu sehen. Bei diesem Prozess sind für die Aktivitäten mit der topologischen ID 1 und 3 die Benutzergruppe mit der Kennung A zuständig. Für die anderen beiden Aktivitäten wird die Benutzergruppe mit der Kennung B zugewiesen. Bei der Erstellung der Schnittmengen resultiert, dass die Aktivität mit der topologischen ID 4 keine korrekte Überlappung für eine Ausführung besitzt, allerdings ist eine kritische Ausführung möglich.

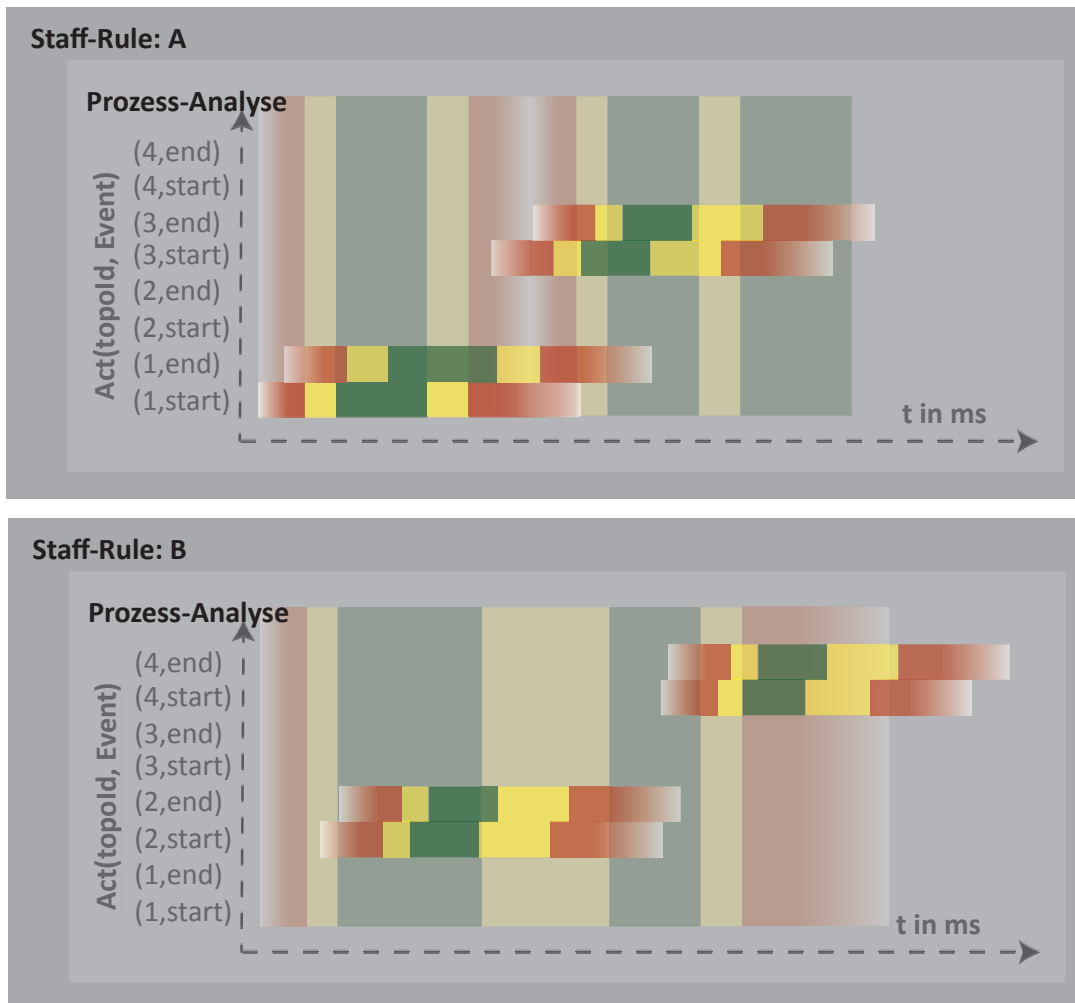


Abbildung 5.8: Online-Analyse und Vorhersage bei Instanz-Start

Dieses Vorhersagen von Zuständen zu Beginn der Instanz und Bestimmung der Zustände der Prozesselemente kann für die Optimierung und Fehlerkorrektur im Scheduler des PMS verwendet werden. Dieses rückwirkende Monitoring wird im letzten Abschnitt dieses Kapitels behandelt.

5.4 Rückwirkendes Monitoring

Ein klassisches Prozess-Managementsystem besteht aus mehreren Komponenten und Managern. Für das rückwirkende Monitoring ist zu Beginn nur der Scheduler interessant. Dieser Scheduler verwaltet und optimiert die Worklists (Listen mit ausführbare Aktivitäten) der einzelnen Benutzer und berechnet die Ausführungsreihenfolge der Aktivitäten. Dieser Scheduler kann durch das Monitoring des vorigen Abschnitts ausgelöst und mit Informationen über eine laufende Instanz und deren Elemente benachrichtigt werden (siehe Abbildung 5.9). Durch diese Information bezüglich der Korrektheit der Instanz kann der Scheduler eine Anpassung und eine Priorisierung von nicht korrekten Instanzen und Aktivitäten vornehmen. Das ist die einfachste Möglichkeit eine Instanz von einem kritischen Zustand in einen korrekten zu überführen. Des weiteren existieren komplexere Verfahren aus dem Bereich der Optimierung [Zuk13], welche allerdings nicht Teil dieser Arbeit sind.

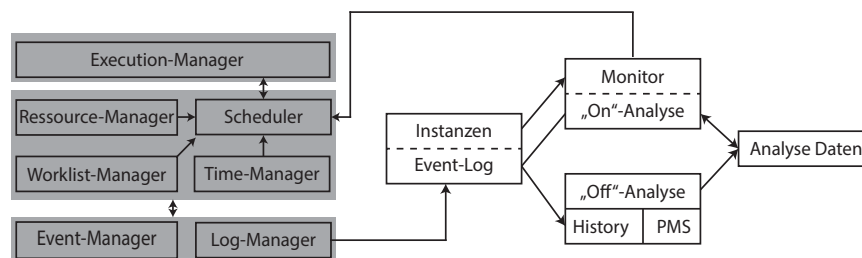


Abbildung 5.9: Aufbau des rückwirkenden Monitorings

Nach der Beschreibung des Monitors, der Vorhersage und des rückwirkenden Monitoring wird im nächsten Kapitel eine Machbarkeitsanalyse für den Monitor und der offline Analyse-Komponente vorgestellt.

6

Prototyp und Implementierung

Nach der Beschreibung der Analyse, des Monitorings, der Vorhersage und der Darstellung von zeitbehafteten Geschäftsprozessen wird in diesem Kapitel eine Machbarkeitsanalyse vorgestellt. Diese Machbarkeitsanalyse enthält folgende Komponenten: das Prozess-Managementsystem *AristaFlow* [DRRM⁺09], die ETL-Komponente der *MSQL Datenbank* für den ersten Transformationsschritt, einer Java-Applikation für den zweiten Transformationsschritt und das Loadings des ETL-Prozesses, dem Monitor und den Prozesssimulator. Diese Komponenten werden im folgenden Kapitel vorgestellt.

6.1 Aufbau des Prototyps

In den vorigen Kapiteln wurden Verfahren und Grundlagen für das Monitoring und die Analyse von zeitbehafteten Geschäftsprozessen vorgestellt. Diese Verfahren wurden

6 Prototyp und Implementierung

teils in einer Machbarkeitsanalyse umgesetzt und untersucht. Hierbei stellte sich ein Zusammenhang und Abhängigkeiten zwischen den einzelnen Komponenten heraus. Das PMS und der Simulator erzeugt während der Ausführung von Instanzen die Einträge im Event-Log. Die Analyse, welche den ETL-Prozess beinhaltet, nutzt diese Informationen aus dem PMS und den Event-Logs für die Erzeugung von den Analyse-Daten. Wie in Abbildung 6.1 zu sehen, werden diese Analyse-Daten dem Monitoring und der Darstellung der Prozesse übergeben. Die Daten können in diesen Komponenten für die Überwachung und Erkennung von kritischen Momenten in Prozessen wiederverwendet werden.

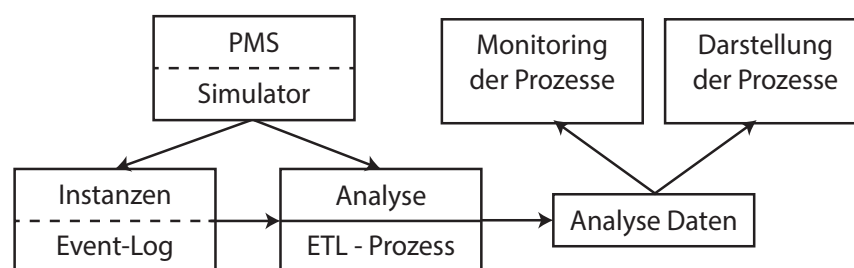


Abbildung 6.1: Aufbau des Prototyps

6.2 Simulator

Basierend auf einem klassischen PMS wurde ein Prototyp eines Simulators für Prozesse im Rahmen eines Projektes entwickelt und implementiert. Dieser Simulator besitzt die typischen Komponenten eines Prozess-Managementsystem (siehe Abbildung 5.1). Jedoch ist dieser Simulator nur für fehlerfreie Ausführungen ausgelegt worden. Das heißt, dieser Simulator kann keine fehlerhafte Dauern von Aktivitäten erzeugen. Um solche kritische und fehlerhafte Dauern zu simulieren, wurde der TimeManager des Simulators erweitert. Dieser Erweiterung erzeugt zu einer Wahrscheinlichkeit von 70% eine korrekte Dauer einer Aktivität, zu 20% eine kritische und zu 10% eine fehlerhafte. Durch diese Möglichkeiten kann der Monitor, welcher in den nächsten Abschnitten vorgestellt wird, in Echtzeit simulierte Prozesse überwachen.

6.3 ETL Prozess - MSOL - JAVA

Wie in Kapitel 4.1 erläutert, müssen für die Darstellung und das Monitoring die Daten durch einen ETL-Prozess aufbereitet, transformiert und angepasst werden. Der erste Schritt der Extraktion der Daten wird im SQL Server Data Tool des MSOL-Servers der Firma Microsoft umgesetzt. Diese Umsetzung findet in einem SSIS-Projekt (SQL Server Integration Service) statt. In den Konnektoren (siehe Abbildung 6.2 Markierungen 1) der Datenquellen werden fehlerhafte und unvollständige Daten entfernt und für die Transformation aufbereitet. Hierbei sind die Datentypen erkannt und dem Arbeitsbereich übergeben worden. Diese Daten werden durch eine Sortierung und Verbindung für die Weiterverarbeitung zusammengefügt (siehe Markierungen 2).

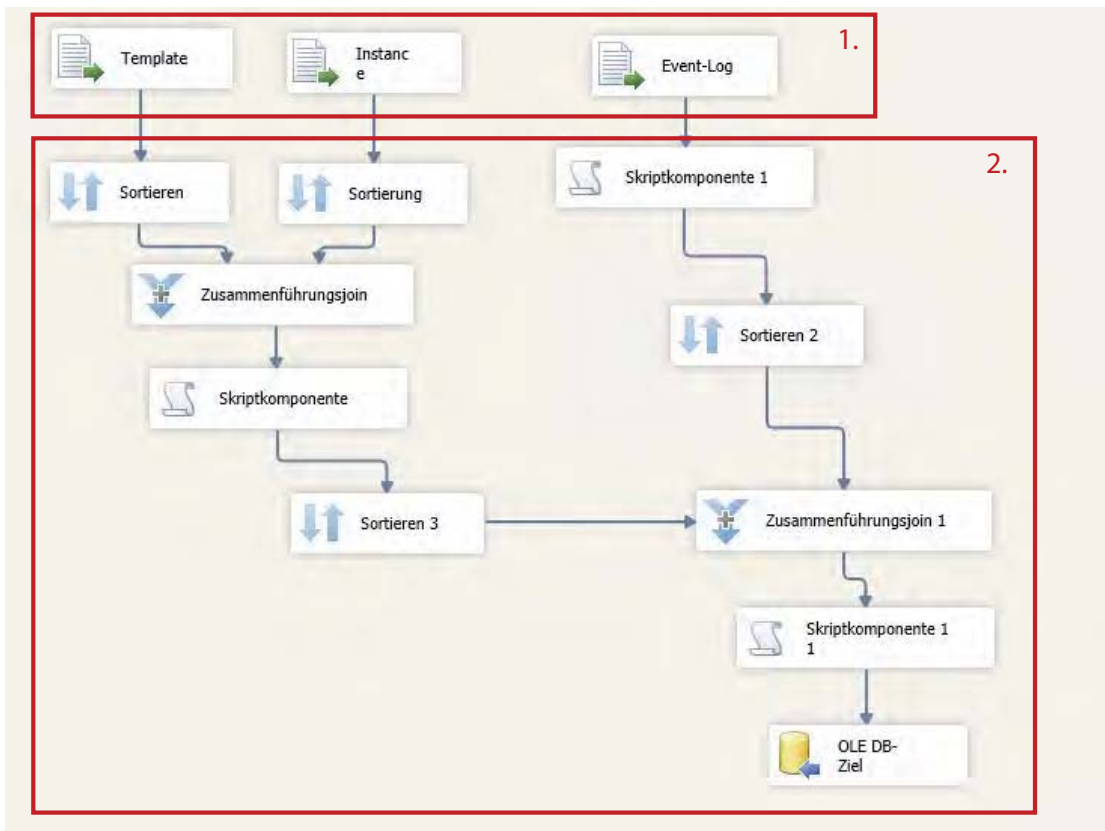


Abbildung 6.2: Integration Service im SQL Server Data Tool

6 Prototyp und Implementierung

Diese Daten und ihre Zeitstempel können anschließend durch eine SQL-Sicht im SQL Server Management Studio zum Verhältnis des Instanzbeginns angepasst und durch die Informationen aus dem PMS erweitert werden (siehe Abbildung 6.3). Nach diesen Anpassungen können die Analyse-Möglichkeiten aus Kapitel 4.2, 4.3 und 4.4 angewandt werden. Dieser zweite Schritt der Transformation wird im nächsten Abschnitt erläutert.

```
SELECT a.nodeid, a.iteration, a.statechange, a.templatename, a.nodename,
       a.timeOut - b.timeOut AS t, a.nodeTopoId, a.templatename, b.templatename
FROM   dbo.ExcelImp1 AS a INNER JOIN dbo.ExcelImp1 AS b
       ON a.templatename = b.templatename AND a.instancelogid = b.instancelogid
WHERE  (b.nodeid = 0) AND (a.iteration = 0) AND (b.iteration = 0)
       AND a.templatename like 'Sim'
Order By(a.nodeTopoId)
GO
```

Abbildung 6.3: Berechnung der Zeitstempel zum Instanz-Start

6.4 Analyse Komponente - Zweiter Transformationsschritt

Wie in Kapitel 5 beschrieben, benötigt die Monitoring-Komponente die Analyse-Daten des Verfahrens aus Kapitel 4, wobei möglichst alle Prozesselemente semantisch korrekt überwacht werden müssen (siehe Kapitel 1 Anforderungen). Diese korrekte Überwachung ist dabei ausschließlich durch das Verfahren aus Abschnitt 4.4 möglich. Diese offline Analyse der zeitbehafteten Geschäftsprozesse wurde in einer Java-Applikation umgesetzt. Diese Java-Applikation erzeugt die Analyse-Daten und entspricht der Loading-Komponente des ETL-Prozesses. Die Daten des Loadings sind in Abbildung 6.4 beschrieben und können für die Monitoring- und Darstellungs-Komponente verwendet werden.

Neben den Verfahren für die Analyse von zeitbehafteten Geschäftsprozessen wurde in Kapitel 4 die Analyse für Ausführungszeiträume vorgestellt. Hierbei wurde die Ausführungszeiten der Aktivitäten nach ihrem Tag (siehe Analyse-Datei in Abbildung 6.5) und je nach Granularität der Stunden, Minuten oder Sekunden einsortiert. Durch diese Sor-

6.5 Monitoring Komponente

| templateName | nodeid | iteration | branch | state | redLow | yellowLow | yellowHigh | redHigh | size | |
|--------------|--------|-----------|----------|-------|--------|-----------|------------|---------|---------|-----|
| | 0 | 3 | 0 [0, 1] | | 16 | 294001 | 294001 | 1176001 | 1836001 | 363 |
| | 0 | 4 | 0 [0, 1] | | 16 | 102000 | 102000 | 408000 | 828000 | 353 |
| | 0 | 5 | 0 [0, 1] | | 16 | 403200 | 403200 | 1612800 | 2620800 | 333 |
| | 0 | 6 | 0 [0, 1] | | 16 | 331200 | 331200 | 1324800 | 2332800 | 320 |

Abbildung 6.4: Loading-Daten der Boxplots

tierung lassen sich beliebige Ausführungszeiträume bestimmen und für die Vorhersage verwenden.

| nodeid | mon | tue | wed | thu | fri | sat | sun | Ingesamt |
|----------|------------|------------|------------|------------|------------|-----------|------------|----------|
| Anzahl | 2626 | 2203 | 2325 | 3107 | 2519 | 318 | 293 | 13391 |
| Prozent: | 19,6101859 | 16,4513479 | 17,3624076 | 23,2021507 | 18,8111418 | 2,3747293 | 2,18803674 | |

Abbildung 6.5: Loading-Daten der Ausführungszeiträume

Im letzten Abschnitt werden die offline und online Monitoring-Komponente vorgestellt, wobei die online Analyse im vorliegenden Monitor nicht implementiert ist.

6.5 Monitoring Komponente

Wie im Abschnitt 6.1 gezeigt, besteht der Prototyp des Monitorings aus zwei Bereichen. Diese Bereiche sind in das offline und online Monitoring eingeteilt. Dieses Monitoring wurde in einer Java-Applikation umgesetzt und visualisiert. Die Visualisierung des offline und online Monitoring ist in Abbildung 6.6 zu erkennen. Die Dauern von Aktivitäten wurde durch die Informationen des Analyse-Verfahrens bewertet, wobei in diesem Beispiel einer Instanz Node 30 als kritisch, Node 28 als fehlerhafte und die Nodes 0, 2, 29 und die laufende Node 34 als korrekt eingestuft wurden. Die Nodes 36 und 35 sind in einem abgewählten Entscheidungs-Zweig. Der Fortschritt, welcher durch das offline Monitoring bewertet wird, bewertet die Instanz als korrekt (grün). Nicht bewertete Prozesselemente wurden hellblau gefärbt.

Durch diese Machbarkeitsanalyse wurde gezeigt, dass die Analyse- und das Monitoring-Verfahren aus Kapitel 4.4 & 5 umgesetzt werden kann. Die Analyse erkennt Entschei-

7

Zusammenfassung

7.1 Fazit

In dieser Arbeit wurde untersucht, wie zeitbehaftete Geschäftsprozesse durch eine Analyse für das Monitoring und Vorhersagen von korrekten Ausführungen in einem PMS aufbereitet werden können. Grundlegend für dieses Verfahren ist das Prozess-Managementsystem *AristaFlow*, welches durch Zeitconstraints und Zeitaspekte in der Prozess-Metasprache *ADEPT2* erweitert wurde. Für diese Verfahren und Komponenten wurden zunächst Anforderungen und Eigenschaften definiert, die unterstützt und abgebildet werden müssen.

Nachdem die Anforderungen und der Rahmen der Arbeit abgesteckt waren, wurden verwandte Verfahren für die Analyse, das Monitoring und für die Vorhersage vorgestellt. Zudem wurde auf Zeitaspekte und Prozesstransformationen eingegangen, welche an-

7 Zusammenfassung

hand ihrer Eigenschaften bewertet und ausgewertet wurden. Die Bewertungen zeigten, dass statistische Verfahren für die Erzeugung der Analyse hervorragend geeignet sind. Insbesondere stellte sich heraus, dass Boxplots, welche ausreißerneutral sind, effizient auf zeitbehaftete Daten angewendet werden können, allerdings wird dazu zunächst eine Transformation und weitere Eigenschaften des Prozesses benötigt.

Im ersten Hauptteil dieser Arbeit, dem Kapitel 4, wurden unterschiedliche Verfahren für die Analyse- und Darstellungs-Möglichkeit vorgestellt. Diese Verfahren sind Teil eines ETL-Prozesses. In diesem ETL-Prozess werden Prozess-Event-Log mit dem Prozessmodell und weitere Informationen aus dem PMS kombiniert. Hierbei wurde auch das Prozessmodell durch eine Transformation in ein Zeitmodell mit weiteren Prozesseigenschaften überführt. Ohne diese Transformation und Abbildung des Prozesses wären nicht alle Anforderungen an Analyse, Monitoring und Vorhersage abgedeckt. Durch solche Transformation können Schleifen, Wiederholungen, Entscheidungen und Sequenzen in beliebiger Variation analysiert und überwacht werden.

Nach der Vorstellung der Analyse- und Darstellungs-Möglichkeit wurde im nächsten Hauptteil das Monitoring und die Vorhersage der zeitbehafteten Geschäftsprozesse behandelt. Dieses Monitoring und die Vorhersage basiert auf den Analyse-Verfahren des vorigen Kapitels der Boxplots mit der Prozesstransformation und weiteren Prozesseigenschaften. Durch diese Analyse-Daten können Dauern von Aktivitäten und Abstände zwischen Aktivitäten in Echtzeit überwacht und durch eine Bewertungsgrundlage einem Zustand zugeordnet werden. Diese Zustände sind je nach Ausführbarkeit dem Umfeld eines Benutzers angepasst und basiert der farblichen Einteilung einer Ampel (*grün/korrekt*, *gelb/kritisch* und *rot/fehlerhaft*). Nach der Einteilung kann der Fortschritt der Instanz gemessen und durch die Zustände bewertet werden, wobei das Monitoring-Verfahren in online und offline Monitoring unterteilt wird. Das online Monitoring ist für die Echtzeitüberwachung zuständig und das offline Monitoring für die Bewertung des Fortschrittes beim Auftreten von Ereignisse. Integriert in diesem Monitor ist die online Analyse, welche die Vorhersage von zeitbehafteten Geschäftsprozessen umsetzt. Diese Vorhersage benutzt die Informationen der Analyse-Daten der Boxplots und verbindet sie mit der Analyse der Ausführungszeiträume aus Kapitel 4.2. Durch diese Verbindung können Aussagen über zukünftige Prozesselemente erstellt und vorhergesagt werden. Durch

solche Vorhersagen kann ein rückwirkendes Monitoring betrieben werden, welches im letzten Abschnitt kurz angesprochen wird.

Nach den Analyse-Verfahren, des Monitorings und der Vorhersage wurde zuletzt eine Machbarkeitsanalyse implementiert. Sie zeigt die Realisierbarkeit der Grundideen der Analyse und des Monitorings. Hierbei wurden verschiedene Technologien, wie das PMS *AristaFlow*, der Datenbank-Server mit der ETL-Komponente *MSQL*-Server und ein auf Java- basierender Prozesssimulator verwendet.

Wie diese Arbeit zeigt ist das Monitoring, die Analyse und Vorhersage von zeitbehafteten Geschäftsprozessen möglich. Das Verfahren der Boxplots mit der Prozesstransformation in Kombination mit weiteren Prozesseigenschaften ermöglicht eine einfache Darstellung der Zeitaspekte in einem Prozess und stellt die Analyse-Daten für das Monitoring und der Vorhersage zur Verfügung. Somit können fehlerhafte und kritische Ausführungen in einer Instanz frühzeitig erkannt und behoben werden. Jedoch sind in diesem Verfahren längst nicht alle Elemente eines Prozesses umgesetzt. Besonders existieren zusätzliche Zeitconstraints, die nicht ohne erheblichen Aufwand auf das angewandte Analyse-Verfahren abgebildet werden können. Neben diesen Verfahren der Analyse der Prozesselementen ist das Verfahren für die Bestimmung von Ausführungszeiträume nur für die komplette Benutzergruppe einer Aktivität möglich. Diese Benutzergruppen sollten feiner bestimmt werden können. Das rückwirkende Monitoring stellt eine einfache Methode zur Behebung fehlerhafter Ausführungen dar, allerdings bleibt die Optimierung der Ausführungsreihenfolgen in Verbindung mit den fehlerhaften Prozesselementen offen und kann in weiterführenden Forschung optimiert werden.

7.2 Ausblick

Wie im vorigen Abschnitt erwähnt wurden nicht alle Eigenschaften von zeitbehafteten Geschäftsprozessen umgesetzt. Diese Eigenschaften, die Ereignisbasiert von einer externen Anwendung geschaltet werden oder fixe Startpunkte von Prozesselementen besitzen, sind durch die Analyse und den Monitor schwer zu erkennen. Daher ist

7 Zusammenfassung

wünschenswert, wenn weitere Zeitaspekte aus der Tabelle 2.3 analysiert, dargestellt und überwacht werden können.

Nach der Betrachtung des Monitorings und die mögliche Erkennung von fehlerhaften Instanzen stellt sich im Weiteren die Frage, weshalb die entsprechende Instanz und Prozesselemente nicht korrekt ablief. Daher sollte eine Generierung von zeitbehafteten Regeln in Geschäftsprozessen für die Fehlererkennung entwickelt werden. Neben diesen Regeln in einer Instanz eines Geschäftsprozesses müssen Abhängigkeiten zwischen parallel ablaufenden Instanzen bewertet und überwacht werden können. Hierbei spielt die Organisationsstruktur aller Benutzer in einem Unternehmen eine wichtige Rolle.

Neben der Erkennung der Fehler durch den Monitor und der Vorhersage müssen kritische und fehlerhafte Instanzen wieder korrigiert werden. Dies wurde in Kapitel 5 des rückwirkende Monitoring angesprochen, jedoch ist dieses Verfahren der Priorisierung nicht effizient und Teils nicht zielführend. Daher müssen Optimierungs-Verfahren in Kombination mit der Fehlerkorrektur einer Instanz in einem PMS angesprochen werden.

Zuletzt sollten die Verfahren des Monitorings, der Analyse, die Darstellung und Vorhersage von dem Prozess-Managementsystem unabhängig sein. Das heißt, dass Prozesse aus dem Prozess-Event-Log selbst erstellt und entdeckt werden. An diese entdeckte Prozesse, kann die Darstellung der Verfahren aus dieser Arbeit annotiert werden. Zudem muss die Machbarkeitsanalyse und der Prototyp weiter entwickelt und mit allen vorgestellten Verfahren ausgerüstet werden. Abschließend lässt sich feststellen, dass durch das Verfahren der Boxplots in Kombination und Transformation von weiteren Prozesseigenschaften die Anforderungen der Analyse erfolgreich abgebildet und somit der Grundstein für ein weites Feld der Optimierung, Darstellung, Fehlerkorrektur, Vorhersage und Überwachung gelegt werden konnte. Diese Verfahren ermöglichen dem Benutzer eine einfache Orientierung über den Zustand einer Instanz und weisen ihn in Echtzeit auf bestehende Probleme im Prozess hin.

Abbildungsverzeichnis

| | | |
|------|--|----|
| 2.1 | Einfacher Prozess in ADEPT | 6 |
| 2.2 | Basiselemente des ADEPT-Metamodells | 8 |
| 2.3 | Pattern Kategorie [LWR10] | 8 |
| 2.4 | Aktivitäten- und Prozessdauer | 9 |
| 2.5 | Prozess mit Zeitkanten in ADEPT | 10 |
| 2.6 | Weiderholungen bzw. Cycles [LWR10] | 12 |
| 2.7 | ADEPT Knotenzustände mit Zeiterweiterung (in Anlehnung an [Rei00]) | 14 |
| 3.1 | Boxplot, Antennen und Ausreißer | 18 |
| 3.2 | Unterschiedliche Histogramme mit gleichem Datensatz | 19 |
| 3.3 | Clustering von Messreihen [KP98] | 21 |
| 3.4 | Verteilungen auf zeitbehafteten Messreihen [RSW12] | 23 |
| 3.5 | Einfache Berechnung und Analysen der Zeitaspekte [ASS11] | 24 |
| 3.6 | Zusammenspiel der Komponenten [BDDW07] | 25 |
| 3.7 | Prozess in Verbindung mit Events [HKRS12] | 26 |
| 3.8 | Verschiedene Event-Typen [BDDW07] | 27 |
| 3.9 | Abstände zum Schluss [CP99] | 28 |
| 3.10 | Transformation von zeitbehafteten Prozessen [LR14] | 29 |
| 3.11 | Aktivität mit Mapping auf CSTNU [LPCR13] | 30 |
| 4.1 | ETL-Prozess | 34 |
| 4.2 | Erstellung des zeitbasierten Histogramms | 37 |
| 4.3 | Zusammenführen von Klassen | 37 |

Abbildungsverzeichnis

| | | |
|------|--|----|
| 4.4 | Ausreißerproblem bei Histogrammen mit Eliminierung | 38 |
| 4.5 | Prozess mit Histogramm Annotationen | 40 |
| 4.6 | Optimales und verschobenes Boxplot | 42 |
| 4.7 | Prozess mit Boxplot Annotationen | 44 |
| 4.8 | Prozess mit Entscheidungen und Schleifen | 45 |
| 4.9 | Problematische Abstände und Dauern | 46 |
| 4.10 | Transformation des Prozessgraphen und der Instanzen | 49 |
| 4.11 | Berechnung der Dauer einer Aktivität | 50 |
| 4.12 | Berechnung des Abstands zwischen Aktivitäten | 50 |
| 4.13 | Schleifen und in Verbindung mit den BranchIDs | 51 |
| 4.14 | Darstellung der Zusammenführung von Boxplots | 52 |
| 4.15 | Analyse des Ausführungszeitraumes | 53 |
| | | |
| 5.1 | Zusammenspiel von Analyse, Daten-Logs und des Monitors | 57 |
| 5.2 | Aufbau des Online Monitorings | 58 |
| 5.3 | Aufbau des Offline Monitorings | 58 |
| 5.4 | Boxplot mit Bewertung | 59 |
| 5.5 | Monitoring von Entscheidungen | 61 |
| 5.6 | Laufende Instanz mit Bewertungen | 62 |
| 5.7 | Online und Offline Monitoring | 64 |
| 5.8 | Online-Analyse und Vorhersage bei Instanz-Start | 65 |
| 5.9 | Aufbau des rückwirkenden Monitorings | 66 |
| | | |
| 6.1 | Aufbau des Prototyps | 68 |
| 6.2 | Integration Service im SQL Server Data Tool | 69 |
| 6.3 | Berechnung der Zeitstempel zum Instanz-Start | 70 |
| 6.4 | Loading-Daten der Boxplots | 71 |
| 6.5 | Loading-Daten der Ausführungszeiträume | 71 |
| 6.6 | View des Monitors | 72 |

Literaturverzeichnis

- [ASS11] AALST, Wil M. d. ; SCHONENBERG, MH ; SONG, Minseok: Time prediction based on process mining. In: *Information Systems* (2011)
- [BDDW07] BARROS, Alistair ; DECKER, Gero ; DUMAS, Marlon ; WEBER, Franz: Correlation Patterns in Service-Oriented Architectures. In: DWYER, MatthewB. (Hrsg.) ; LOPES, Antónia (Hrsg.): *Fundamental Approaches to Software Engineering* Bd. 4422. Springer Berlin Heidelberg, 2007
- [BH06] BHATTACHARYA, Arindam ; HEMERLING, Jim: Organizing for Global Advantage in China, India, and Other Rapidly Developing Economies. In: *The Boston Consulting Group* (2006)
- [BSS00] BARTEL, Wolfgang ; SCHWARZ, Stefan ; STRASSER, Gerhard: Der ETL-Prozess des Data Warehousing. In: *Data Warehousing Strategie*. Springer, 2000
- [Bul13] BULLINGER, Hans-Jörg: *Wege aus der Krise: Geschäftsprozessoptimierung und Informationslogistik*. Bd. 41. Springer-Verlag, 2013
- [Cha83] CHAMBERS, John M.: *Graphical methods for data analysis*. 1983
- [CP99] CHEN, Xiaodong ; PETROUNIAS, Ilias: Mining temporal features in association rules. In: *Principles of Data Mining and Knowledge Discovery*. Springer, 1999
- [DRRM⁺09] DADAM, Peter ; REICHERT, Manfred ; RINDERLE-MA, Stefanie ; GÖSER, Kevin ; KREHER, Ulrich ; JURISCH, Martin: Von ADEPT zur AristaFlow BPM

Literaturverzeichnis

- Suite-Eine Vision wird Realität:"Correctness by Construction und flexible, robuste Ausführung von Unternehmensprozessen. (2009)
- [GCC⁺04] GRIGORI, Daniela ; CASATI, Fabio ; CASTELLANOS, Malu ; DAYAL, Umeswar ; SAYAL, Mehmet ; SHAN, Ming-Chien: Business process intelligence. (2004)
- [GHS95] GEORGAKOPOULOS, Diimitrios ; HORNICK, Mark ; SHETH, Amit: An overview of workflow management: From process modeling to workflow automation infrastructure. (1995)
- [HKRS12] HERZBERG, Nico ; KUNZE, Matthias ; ROGGE-SOLTI, Andreas: Towards Process Evaluation in Non-Automated Process Execution Environments. In: *ZEUS*, 2012
- [JB96] JABLONSKI, Stefan ; BUSSLER, Christoph: *Workflow Management: Modeling Concepts, Architecture and Implementation*. International Thomson Computer Press, 1996
- [KP98] KEOGH, Eamonn J. ; PAZZANI, Michael J.: An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback., 1998
- [LM08] LAUE, Ralf ; MENDLING, Jan: *The impact of structuredness on error probability of process models*. Springer, 2008
- [LPCR13] LANZ, Andreas ; POSENATO, Roberto ; COMBI, Carlo ; REICHERT, Manfred: Controllability of time-aware processes at run time Springer, 2013
- [LR14] LANZ, Andreas ; REICHERT, Manfred: Analyzing the Impact of Process Change Operations on Time-Aware Processes. (2014)
- [LWR10] LANZ, Andreas ; WEBER, Barbara ; REICHERT, Manfred: Workflow Time Patterns for Process-aware Information Systems. Hammamet, Tunisia : Springer Berlin / Heidelberg, 2010
- [McC02] MCCOY, David W.: Business Activity Monitoring: Calm Before the Storm. Gartner. In: *Inc., Apr* (2002)

- [RA08] ROZINAT, Anne ; AALST, Wil M. d.: Conformance checking of processes based on monitoring real behavior. In: *Information Systems* (2008)
- [RCWR04] REICHERT, Manfred ; CATRINESCU-WIEDEMUTH, Ursula ; RINDERLE, Stefanie: Evolution von Zugriffsregelungen in Informationssystemen. (2004)
- [Rei00] REICHERT, Manfred: Dynamische Ablaufänderungen in Workflow-Management-Systemen, University Ulm, 2000
- [RSW12] ROGGE-SOLTI, Andreas ; WESKE, Mathias: Enabling probabilistic process monitoring in non-automated environments. In: *Enterprise, Business-Process and Information Systems Modeling*. Springer, 2012
- [SCO79] SCOTT, DAVID W.: On optimal and data-based histograms. In: *Biometrika* (1979)
- [SS08] SCHMELZER, Hermann J. ; SESSELMANN, Wolfgang: Geschäftsprozessmanagement in der Praxis. In: *Kunden zufrieden* (2008)
- [Tuk77] TUKEY, John W.: Exploratory data analysis. In: *Reading, Ma* 231 (1977)
- [vv02] VAN DER AALST, Wil ; VAN HEE, Kees: *Workflow Management: Models, Methods, and Systems*. The MIT Press, 2002
- [Zuk13] ZUKERMAN, Moshe: Introduction to queueing theory and stochastic teletraffic models. In: *arXiv preprint arXiv:1307.2968* (2013)

Name: MARTIN SOMMER

Matrikelnummer: 643177

Erklärung

Ich erkläre, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ulm, den

MARTIN SOMMER