

Abschlussbericht zu SoftProc

Zuwendungsempfänger: PITSS GmbH Universität Ulm	Förderkennzeichen: 01IS20027A-B
Vorhabenbezeichnung: SoftProc - Entwicklung von Verfahren zur retrograden Prozessanalyse aus einer laufenden Applikation	
Laufzeit des Vorhabens: 01.04.2021 bis 31.03.2023	

Inhaltsverzeichnis

1. Ausgangsbasis	1
1.1. Ursprüngliche Aufgabenstellung	1
1.2. Zielsetzung	1
1.3. Wissenschaftlicher und technischer Startpunkt	2
1.4. Zusammenfassung der wesentlichen Projektergebnisse	2
1.4.1. PITSS GmbH	2
1.4.2. Universität Ulm.....	3
2. Detaillierte Ausführung	3
2.1. Ausführliche Darstellung der durchgeführten Arbeiten im Vergleich zur Vorhabensbeschreibung	3
3. Wichtigsten Positionen des zahlenmäßigen Nachweises	15
3.1. PITSS GmbH	15
3.2. Universität Ulm	16
4. Notwendigkeit und Angemessenheit der Projektarbeiten.....	17
4.1. PITSS GmbH	17
4.2. Universität Ulm	17
5. Voraussichtlicher Nutzen, insbesondere Verwertbarkeit.....	18
5.1. PITSS GmbH	18
5.2. Universität Ulm	20
6. Fortschritte in Wissenschaft und Technik während der Projektlaufzeit.....	20
7. Erfolgte und geplante Veröffentlichungen.....	21

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS20027A-B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim der Autorin/beim Autor.

1. Ausgangsbasis

1.1. Ursprüngliche Aufgabenstellung

Die Implementierung von Softwareanwendungen adressiert die Bedürfnisse von Benutzern, Anwendungsfällen und Geschäftsprozessen. Gängige Software-Systeme sind jedoch mehrheitlich nicht darauf ausgelegt, prozessbezogene Daten hoher Qualität bereitzustellen, um mit modernen Process Mining Werkzeugen umfassende Prozessanalysen und -visualisierungen vorzunehmen. Relevante Fragestellungen sind z.B., wie der von der Software implementierte Prozess aussieht (Process Discovery) und wie stark dieser vom definierten Soll-Prozess abweicht (Conformance Checking).

Da in den meisten Fällen die Individualsoftwarelösung nicht zwingend eine Workflow- bzw. Prozess-Engine im Hintergrund nutzt, die mit wenigen Erweiterungen hilfreiche Audit-Daten zu liefern vermag, bleibt vielfach nur der Blick auf die Veränderungen im Datenmodell. Diese sind allerdings mit viel Zeitverlust durch Interviews, Analysen und Vermutungen verbunden.

1.2. Zielsetzung

Ziel des Vorhabens war die Entwicklung softwaregestützter Verfahren, die Anwender bei der Abarbeitung eines Prozesses mit der Software „beobachtet“ und diese Informationen feingranular in einen Event-stream bzw. ein Event Log schreibt. Mit der resultierenden Online-Prozessanalyse werden aus einer laufenden Applikation dedizierte Prozesse retrograd ermittelt und die Verwendung der Anwendung über unterschiedliche Anwender, Standorte oder Zeitpunkte dokumentiert.

Untersucht wurden verschiedene Aufzeichnungsvarianten für Produktivsysteme, die über Konfigurationstabellen gesteuert werden können:

(1) Bei der dedizierten Aufzeichnung (*Dedicated Recording*) werden bekannte Prozesse dokumentiert und abhängige Programmkomponenten zugeordnet, wodurch eine Neustrukturierung der Applikation vorgenommen werden kann. Der Anwender bestimmt dabei durch eine Tastenkombination selbstständig den Start und das Ende des aufzuzeichnenden Prozesses und grenzt so exakt die den Prozess ausmachende Funktionalität ab.

(2) Beim *Silent Recording* wird von der ersten Anmeldung des Benutzers bis zum Schließen der Applikation die gesamte Verwendung der Applikation protokolliert. Im Detail kann entschieden werden, ob das für alle Sessions erfolgen soll oder nur für ausgewählte User-Sessions (z.B. Sessions der Fachbereichs-User). Ferner kann konfiguriert werden, welche Detailinformationen (z.B. zur Einhaltung von Datenschutzvorgaben) gespeichert werden sollen.

Mit dem *Silent Recording* in der **Testumgebung** wurde das Performance-Verhalten der Applikation nach Indizierung des Programmcodes getestet. Gerade bei rekursiver Programmierung oder sehr hoher Modularisierung, zeigte sich deutlich der Overhead für das Logging der Programm-Units. In diesen Fällen wurde für offensichtliche Stellen innerhalb der Applikation eine Begrenzung der Indizierung untersucht. In Extremfällen wurde das Logging auf User-Interaktionen (z.B. Key- und Button-Trigger) aus der Bedienoberfläche reduziert, um den eigentlichen Eventstart zu dokumentieren.

In der **Produktivumgebung** werden alle Anpassungen aus dem *Dedicated* und *Silent Recording* umgesetzt, so dass Aufzeichnungen prozessbezogener Ereignisse hier kontrolliert ausgewählt bzw. aktiviert werden können.

1.3. Wissenschaftlicher und technischer Startpunkt

Die gängigen Ansätze, um zu Modellen für Geschäftsprozesse zu gelangen, haben gemeinsam, dass die zu analysierenden Software-Systeme einen möglichst akkuraten Eventstream an ein Process Mining Tool übergeben wollen. In 2016 hat die IEEE das Standardformat XES (Extensible Event Stream) für Eventstreams definiert.¹ Dieser Standard wird von gängigen Process Mining Werkzeugen unterstützt, er erlaubt u.a. eine freie Erweiterbarkeit mittels eigener Extensions, um Spezifika des Unternehmens bzw. seiner Geschäftsprozesse abzubilden, ebenso wie Spezifika der Software-Systeme und die Eventstreams an das Process Mining Tool zu übergeben. Für die großen Standardsysteme mit konstanten Datenmodellen (z.B. SAP, Microsoft Dynamics, Oracle E-Business Suite) gibt es bereits vorkonfigurierte Exportstrecken, die z.B. für einen klassischen Beschaffungsprozess alle notwendigen Eventstream-Informationen aus der Applikation extrahieren. Jedoch gibt es keine simplen Exportstrecken für Standard-Anwendungen oder Branchenlösungen, geschweige denn für Individualsoftware. Da in den meisten Fällen die Individualsoftwarelösung nicht zwingend eine Workflow- bzw. Prozess-Engine im Hintergrund nutzt, die mit wenigen Erweiterungen hilfreiche Audit-Daten zu liefern vermag, bleibt vielfach nur der Blick auf die Veränderungen im Datenmodell. Diese sind allerdings, wie bereits oben erörtert, mit viel Zeitverlust durch Interviews, Analysen und Vermutungen verbunden.

1.4. Zusammenfassung der wesentlichen Projektergebnisse

Im Rahmen der Forschungs- und Entwicklungsarbeiten wurde sichtbar, dass die größte Innovation innerhalb der Arbeitspakete in der Identifikation der Prozesse auf Basis von User-Interaktionen, Identifikation des "realen" Datenmodells und dem DataActivityLog liegt. Sofern daraus erst eine Prozessbeschreibung mit BPMN oder PHILharmonicFlows erstellt werden kann, waren die weiteren Innovationen im Bereich Konformität und Compliance nur noch geringfügig.

Aus diesem Grund lag im weiteren Projektverlauf der Fokus deutlich stärker auf einer hohen Qualität bei der semiautomatischen Erkennung der in der Anwendung implementierten Businessprozesse.

1.4.1. PITSS GmbH

Die wesentlichen Ergebnisse auf Seiten der PITSS lassen sich wie folgt benennen:

- Analyse über Datenmodell
Änderungen am Datenmodell, die durch die Benutzerinteraktionen ausgelöst wurden, kann eine semantische Bedeutung im Umfeld der zu identifizierenden Geschäftsprozesse zugeordnet werden.
- Nutzbarkeit von Applikation und Database-Server Logfiles
Mit der Analyse von Logfiles des Applikation Server und der Datenbank kann auch ohne invasives Tracken der Benutzerinteraktionen nahezu der gleiche Detailgrad an Informationen ermittelt werden, wie mit einer Änderung der Software zur Integration der Tracking-Informationen.
- Erweiterung von Eventstreams für Aktionen, die direkt aus dem Framework erfolgen
Ein Entwickler kann ohne Implementierung von PL/SQL DML-Statements innerhalb der Applikation nutzen. Hierzu wurde ein Verfahren entwickelt, um die zugehörige User-Interaktion und die Änderungen an den Applikationsdaten zuordnen zu können.

- Fehlerfreie Zuordnung von User-Interaktionen zu Modifikationen im Datenmodell

- Verdichtung von feingranularen Mikro-Events zu Business Processen.

Durch Eventlog-Abstraction konnten trotz der Vielzahl der feingranularen Events, die in einem einzelnen Prozess auftauchen können, die wesentlichen Schritte innerhalb eines Prozesses ermittelt und visualisiert werden.

Diese Abstraktion gelingt rein auf Basis der Software-Struktur und der Erkennung der Implementierung ohne notwendiges Applikationswissen.

- Transformation der Implementierung in eine natürliche Prozessbeschreibung

Da auch die Implementierung des reinen User-Interface Bestandteil der automatisierten Softwareanalyse waren, können bspw. auch GUI-Beschriftungen für eine automatische Beschreibung des Prozesses als natürlich-sprachliche Prozessbeschreibung benutzt werden.

Alle angesprochenen Ergebnisse und Erkenntnisse werden in das kommende Release des Produktes PITSS.CON einfließen und stehen somit der Kundenbasis als update oder auch für Neukunden als eigenständiges Produkt zur Verfügung.

1.4.2. Universität Ulm

Das Institut für Datenbanken und Informationssysteme (DBIS) forscht seit 1990 in den Bereichen der prozess- und datenorientierten Informationssysteme sowie zu aktueller Datenbank- und Prozess-Management Technologie (Process & Data Science).

Im Fokus der Forschung der Arbeitsgruppe von Prof. Dr. Manfred Reichert steht die Flexibilität prozessorientierter Informationssysteme entlang ihres Lebenszyklus (Analyse / Mining, Design, Konfiguration, Betrieb und Optimierung).

Prof. Reichert, DBIS-Direktor, zählt zu den weltweit führenden und meistzitierten Forschern im Bereich Business Process Management. Mit ADEPT wurde eine innovative Technologie geschaffen und als AristaFlow BPM Suite in die Praxis transferiert, die derzeit weltweit die Sperspitze flexibler und adaptiver Prozessmanagement-Systeme bildet.

Ebenso einschlägig sind seine Arbeiten aus dem PHILharmonicFlows Projekt zu datenobjekt-zentrierten Prozessen, bei dem die Modellierung, Ausführung und Analyse von Prozessen und Daten integriert erfolgen. Daten- objekt- und benutzerzentrierte Prozesse sind für OLTP- und ERP- Anwendungen typisch. Weitere grundlegende Forschungsbeiträge liegen in Bereichen wie Process Mining, Process Analytics, Process Compliance, Prozessqualität, Prozessmodellierung sowie Prozessautomation.

Das DBIS hat in dem Vorhaben die PITSS mit neuesten Erkenntnissen aus der Wissenschaft und Forschung unterstützt und ihre entsprechende Expertise in das Projekt eingebracht. Sie hat ferner die Algorithmen und Verfahren zur Feststellung der Konformität und Compliance implementierter Prozesse mit einem gegebenen Soll-Prozessmodell bzw. mit bestehenden Compliance-Regeln konzipiert und evaluiert. Ferner hat DBIS die besonders originellen Ergebnisse seiner im SoftProc-Projekt geleisteten Forschungsarbeiten hochkarätig publiziert sowie auf 12 internationalen Konferenzen und Workshops präsentiert (vgl. Publikationsliste in AP12).

Detaillierte Ausführung

2. Ausführliche Darstellung der durchgeführten Arbeiten im Vergleich zur Vorhabensbeschreibung

Im Nachfolgenden sollen die wissenschaftlich-technischen Ergebnisse des Projektes und deren gesammelten Erfahrungen nach Arbeitspaketen vorgestellt werden.

Arbeitspaket AP1: Projektmanagement

Als Konsortialführer übernahm die PITSS die Gesamtprojektkoordination für das Vorhaben. Dies umfasste die Planung, Koordination und Durchführung der Projekttreffen sowie die Abstimmungen zwischen den Entwicklungspartnern PITSS und DBIS sowie den (assoziierten) Anwendungspartnern. Darüber hinaus übernahm die PITSS die Durchführung des Projektmanagements mit dem projektinternen Qualitätsmanagement sowie dem projektbegleitenden Berichtswesen und der zugehörigen Dokumentation für den Projektverbund. Dies umfasste insbesondere auch die Kommunikation mit dem Projektträger.

Das DBIS-Institut der Universität Ulm übernahm im Rahmen dieses Arbeitspakets die Projektmanagement-Aufgaben für ihr Teilprojekt und ihre Arbeitspakete (bspw. Einhaltung von Kosten, Leistung, Terminen | Reporting | etc.) und beteiligte sich aktiv an den bilateralen Absprachen und Projekttreffen mit der PITSS. Abschließend unterstützte das DBIS die PITSS bei der Abstimmung und Detailplanung der Entwicklungsaufgaben sowie der Umsetzung einer Feinplanung.

Arbeitspaket AP2: Technische Voruntersuchungen und Anforderungserhebung, Pflichtenheft

AP 2.1: Bestandsaufnahme typischer Prozessumgebungen

Für die Zusammenarbeit der beiden Projektpartner wurde zu Beginn ein Dictionary erstellt, das die wesentlichen Begrifflichkeiten definiert, die für den beidseitigen Austausch notwendig sind. Insbesondere wurden sowohl technische Aspekte der Prozessimplementierung als auch fachliche Begrifflichkeiten der Geschäftsprozessunterstützung eindeutig festgelegt. Hierbei wurde es notwendig, den Prozessbegriff auf mehreren Ebenen (fachlich, technisch) zu definieren und in seinen verschiedenen Ausprägungen zu betrachten. Die am meisten abstrahierte Bezeichnung ist angelehnt an die Literatur und wird als Geschäftsprozess (engl. *Business Process*) bezeichnet.

Ein einzelner Durchlauf eines Prozesses wird als *Case* bzw. Prozessinstanz bezeichnet. Mehrere Cases des gleichen Business Processes können verschiedene *Attribute* besitzen und sind somit als *Varianten* erkannt.

AP 2.2: Erhebung und Formalisierung der Anforderungen

Zur Ermittlung der Anforderungen an die zu entwickelnde Lösung bzw. Technologie wurde ein Ansatz verfolgt, bei dem repräsentative Persona mit ihren typischen Arbeitsaufgaben und Problemstellungen definiert wurden.

Konkret wurden verschiedene Persona beschrieben, die aus Kundensicht die potenziellen Zielgruppen der entwickelten Lösung darstellen könnten. Die Entscheidung fiel auf die weitere Betrachtung von drei repräsentativen Persona, deren Anforderungen, Knowhow, Erwartungshaltung und Aufgabenstellungen im Detail beschrieben wurden. Dabei sind folgende Persona entstanden:

1. CEO: Fokus ist die Erhebung möglichst vieler Informationen für eine fundierte Modernisierung der Applikation

2. Projekt-Manager: Fokus ist die Planungs- und Durchführungssicherheit bei der Umsetzung abgeleiteter Modernisierungsprojekte
3. Software-Entwickler: Fokus ist die Erkennung von Abhängigkeiten der Software-Entwicklung zu den jeweiligen Geschäftsprozessen

Für jede dieser 3 Persona wurde in einer zweiten Runde nach dem „How might we?“-Verfahren eine Aufstellung potenzieller Lösungsansätze für die jeweilige Person erstellt.

AP 2.3: Feinkonzeption, Anforderungskatalog

Basierend auf den Ergebnissen aus AP 2.2 wurden eine Detailabstimmung der Entwicklungsaufgaben zwischen den Kooperationspartnern PITSS und DBIS durchgeführt und Arbeitspakete verfeinert. Die entsprechenden Festlegungen bzw. Vereinbarungen wurden für alle Projektbeteiligten über Confluence bereitgestellt.

Für die Integration von Fremdsystemen, gerade in dem Bereich Process Mining, wurde von Anfang an auf den Standard IEEE 1849-2016 XES zurückgegriffen. Dieser ermöglicht die Übermittlung aller Informationen an jedwede Produkte und erlaubt gleichzeitig den Freiheitsgrad zur Definition eigener Tags innerhalb des XML-Namespaces. Verschiedene Integrationen von Visualisierungstools für BPMN-Notationen wurden evaluiert, aber für den aktuellen Projektbezug verworfen. Im Vordergrund stand die effiziente, qualitativ hochwertige und automatisierte Erhebung der Rohdaten.

In einer GAP-Analyse wurde erhoben, welcher Detailgrad und welche Daten zur Erreichung des Projektziels notwendig waren. Im Wesentlichen stand dabei die Identifikation des primären Geschäftsobjektes, an Hand gezielter Fragestellungen, d.h. wie kann ohne Prozess- und Kontext-Wissen über die Applikation automatisiert erkannt werden, in welchen Tabellen die „wichtigen“ Objekte gespeichert und wie können sie eineindeutig identifiziert werden, im Vordergrund. Dazu musste ein Verfahren entwickelt werden, um das Datenmodell zu analysieren und daraus resultierende Erkenntnisse in die Analyse der Loginformationen einfließen zu lassen. Dabei galt stets zu betrachten, dass die Änderungen keinen Einfluss auf die Gesamtperformance der Applikation hatten.

Arbeitspaket AP3 - Entwicklung von Verfahren zur Code-Analyse von laufenden Anwendungen

AP 3.1 Entwicklung von Verfahren zur automatischen Erfassung von Prozessentitäten in bestehenden Softwareanwendungen (ProcessTracking bzw. Logging)

Gegenstand des Arbeitspakets AP 3.1 ist die Betrachtung einer Prozessentität als Komposition von Mikroprozessen. Hierzu wurden von den Projektpartnern die einzelnen Mikroprozesse durch Betrachtung und Beurteilung von Aktoren wie z.B. Befehlsflächen, Transaktionssteuerungen oder Benutzer-Navigtionen im User Interface identifiziert.

Für die Identifikation von Mikroprozessen, die zu einer Prozessentität gehören, wurden zwei verschiedene Ansätze verfolgt. Zum einen ein transaktionsorientierter und zum anderen ein interaktionsorientierter Ansatz.

Bei dem transaktionsorientierten Ansatz wurde ein Mikroprozess als Aktionskette zwischen zwei Savepoints oder Commits innerhalb der Datenbank betrachtet. Dieser Ansatz exkludiert

zusätzliche Prozessschritte, die nur reinen lesenden Charakter haben und somit für Änderungen und Persistenz der Daten innerhalb der Applikation irrelevant sind. Der Fokus liegt ausschließlich auf den DML-Statements der Datenbank.

Bei dem interaktionsorientierten Ansatz wurde ein Mikroprozess als Aktionskette zwischen zwei „relevanten“ User-Interaktionen betrachtet. Dabei wurde das Modell so konzipiert, dass es durch Konfigurationen die „relevanten“ User Interaktionen selbstständig erkennen konnte.

<pre> 1 declare 2 var as number; 3 begin 4 do_something(var); 5 exception when others then 6 capture_error; 7 end;</pre>	<pre> 1 declare 2 ":A" number; 3 var as number; 4 begin 5 Rec.LogStory(Rec.GetStory, '10.ScreenName.EventName', '*start*', ":A"): 6 do_something(var); 7 Rec.LogStory(Rec.GetStory, '10.ScreenName.EventName', '*exit*', ":A"): 8 exception when others then 9 Rec.LogStory(Rec.GetStory, '10.ScreenName.EventName', '*exception*', ":A"): 10 capture_error; 11 end;</pre>
--	---

Abbildung 1: Codevergleich: Sourcecode ohne Tracker (links) und mit Tracker (rechts)

Die Code-Analyse wurde so entwickelt, dass im weiteren Verlauf Daten für beide Ansätze zur Verfügung standen.

AP 3.2 Entwicklung geeigneter Abstraktionsebenen und Sichten auf erfasste Prozesse

Im Rahmen des Arbeitspakets AP 3.2 wurde ein Datenmodell entwickelt mit dem Prozessschritte, Mikroprozesse, Prozessentitäten und Prozessfamilien abgebildet werden können. Das Datenmodell wurde dabei durchgängig in Form von Graphen abgebildet, was den leichten Einsatz von Graph-Algorithmen, z.B. die Erkennung von sich wiederholenden Komponenten oder die Bildung von Clustern, ermöglichte. Zusätzlich wurden durch strukturelle Abstraktion der Events drei verschiedene Verdichtungen dieser Graphen erzeugt, um auf unterschiedlicher Ebene ein Discovery von Prozessen durchführen zu können.

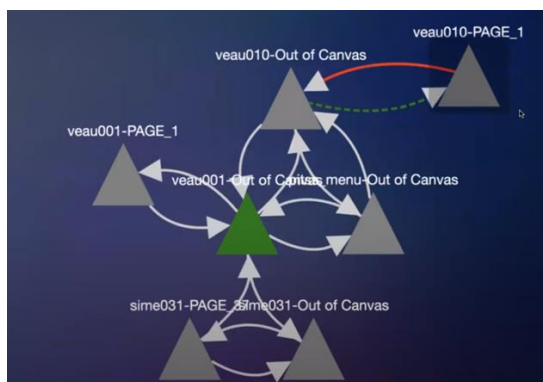
AP 3.3 Untersuchungen zur möglichen Nutzung bzw. Adaption von Machine Learning und Process Mining Verfahren

Mit der Übergabe der Daten aus den verschiedenen Abstraktionsebenen an Process Mining Werkzeuge (im Projektanwendungsfall PhilHarmonic-Flows und PM4Python) konnte ein Discovery auf diesen Ebenen durchgeführt werden. Die erkannten Prozesse wurden im Gespräch mit den Key-Usern verifiziert.

Für eine kontinuierliche Auswertung der Prozesse in einem eventuellen Dashboard wurden alle Teilschritte (Einlesen des DB-Logfiles, Anreicherung der Anwender-Interaktionen, Schreiben der XES-Files, etc.) so konzipiert, dass sie ggfls. als kontinuierliche Jobs regelmäßig ausgeführt werden können.

AP 4.2 Untersuchung von Verfahren und Konzepten zum Dedicated sowie Silent-Recording und Adaption der Konzepte für das Vorhaben

Für die Identifikation der Prozesse innerhalb des Silent-Recording wurde versucht, aus den dedizierten Aufzeichnungen Marker zu extrahieren, die eine eindeutige Zuordnung erlauben. Die Entscheidung wurde getroffen, dass neben dem Start und Ende des Prozesses der User im Nachgang der dedizierten Aufzeichnung die Möglichkeit bekommen sollte, Key-Aktionen markieren zu können. Diese Marker wurden in die Verdichtung des Eventstreams übernommen und können so helfen, die Prozesse innerhalb des Silent-Recording wieder zu erkennen.



Erläuterung zum Bild: Es zeigen sich viele Wechsel zwischen einzelnen Formularen. Das bedeutet, dass der User viel zwischen den Formularen hin und her springt – Optimierungspotenziale für den Prozess werden ersichtlich, um den Ablauf zu vereinfachen bzw. zu verkürzen

Abbildung 3: Beispiel einer Forms Flow – Userverhalten im dedicated Recording

Im gesamten Verlauf des Projektes wurden immer wieder Performance-Analysen durchgeführt, um sicherzustellen, dass das notwendige Logging gerade im Silent-Recording nicht die User-Experience beeinträchtigt. Als Konsequenz wurde ermöglicht, dass gerade rekursive oder hochfrequente Units, die keinen businessprozess-relevanten Code enthalten, vom Aufzeichnen ausgeschlossen wurden.

AP 4.3 Entwicklung und Implementierung der Softwarealgorithmen

Ausgehend von den in AP 4 vorangegangenen Entwicklungsergebnissen wurden diese in Softwaremethoden zur Verarbeitung der Rohdaten und deren Verdichtung zu aussagekräftigen Informationen auf der Ebene von Geschäftsprozessen überführt und implementiert.

Die Softwaremethoden aus diesem Arbeitspaket wurden darüber hinaus kontinuierlich in ein Gesamtdemonstrator integriert und getestet.

Arbeitspaket AP5 - Untersuchungen zur Extraktion von Event-Daten

AP 5.1 Entwicklung von Verfahren, die für Prozesse dazugehörigen Informationen (Metadaten) aus einer Applikation erfassen und in einem Datenmodell abbilden

Die ersten Analysen in diesem Arbeitspaket zeigten sehr schnell, dass nicht davon ausgegangen werden kann, dass eine Legacy-Applikation ein sauber modelliertes ERM aufweist. In einem Modell, in dem jede Tabelle als unabhängiger Datentopf ohne Verknüpfung zu anderen Objekten festgehalten ist, fällt es schwer semantische Entitäten oder gar Business Objekte

automatisiert erkennen zu können. Durch ein vollständiges Parsen der Applikation, inklusive aller implementierten Datenzugriffe aus dem User Interface können aber Verknüpfungen zwischen Objekten in der Persistenz erkannt und durch ihre Häufigkeit bewertet werden. Damit kann ein ERM rekonstruiert werden, welches eine „Verwendung“ des Datenmodells beschreibt. Damit greifen Graphen-Algorithmen zur Clusterung der Objekte und somit einer validen Vorlage für die Erkennung von Business Objekten.

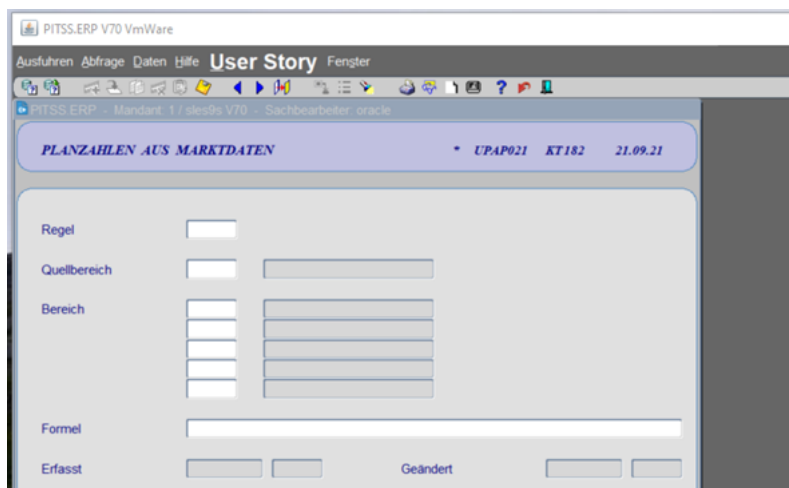


Abbildung 4: Screenshot der genutzten Legacy Application

Das Parsen der Applikation inklusiver aller Objekte, die auch im User Interface modelliert wurden, schaffte ein Grundlage, Objekte, Attribute und auch Benutzer Interaktionen natürlichsprachlich zu bezeichnen. Das Vokabular wurde auf Basis von Feldbezeichnungen, Buttonbeschriftungen, Menüeinträgen, etc. erkannt und für die Prozessbeschreibung aufgegriffen. Sollte die Anwendung als mehrsprachige Applikation entworfen worden sein, so kann am Ende auch die Prozessbeschreibung mehrsprachig zur Verfügung gestellt werden.

AP 5.2 Untersuchung und Adaption von „Abstract Syntax Tree“ (AST)-Konzepten

Für eine bessere Verknüpfung der Benutzer-Interaktionen mit den daraus resultierenden Änderungen im Datenmodell, wurden automatisiert weitere Tracker-Informationen in die Applikation eingeführt. Auf Basis des identifizierten AST (Abstract Syntax Tree) der jeweiligen Programm-Blöcke, wurden alle DML-Statements für eine automatisierte Erkennung eingefasst und die notwendigen primären Attribute in das Eventlog übertragen. Darüber hinaus wurden automatisiert eigene Programm-Blöcke erstellt, die die DML-Statements, die automatisiert vom Framework Oracle Forms und nicht durch eine explizite Programmierung in der Legacy-Applikation ausgeführt werden, in der Prozessaufzeichnung berücksichtigt.

AP 5.3 Implementierung der entwickelten Softwarefunktionen

Die in AP 5.1 und AP 5.2 entwickelten Verfahren wurden in Softwaremethoden umgesetzt und in den Demonstrator integriert.

Arbeitspaket AP6 - Erarbeitung und Evaluation eines geeigneten Datenschutzkonzepts

AP 6.1 Erarbeitung des Datenschutzkonzepts

Gerade für den Bereich des Silent-Recording war sicherzustellen, dass die User-Interaktion inkl. der daraus resultierenden Datenmanipulationen aber auch insbesondere Leistungsdaten nicht auf die angemeldeten Anwender zurückgeführt werden konnten.

Ausgangslage in der Evaluation war die Tatsache, dass ausschließlich existierende Oracle Forms basierende Anwendung Gegenstand des Projektes waren. Innerhalb einer solchen Infrastruktur ist es sehr gebräuchlich, dass sich Anwender mit einem bestehenden Oracle Datenbank-User am System authentifizieren müssen. Diese Art der Legacy-Applikationen unterstützte lange Zeit kein Social-Login oder andere Single-Sign-On Konzepte, die mit bestehenden Directories in Verbindung gebracht werden konnten.

Innerhalb der Oracle Datenbank haben bereits jetzt Datenbank-Administratoren die Möglichkeit, laufende Transaktionen einem angemeldeten Benutzer und einem Terminal zuzuordnen. Ebenfalls ist dem Log-File der Oracle Datenbank entnehmbar, welcher Datenbank-User zu welcher Zeit welche Änderungen durchgeführt hat.

Aufgabe des Datenschutzkonzeptes war es somit, aus den verfügbaren Quellen im Silent-Recording nur die Informationen zu entnehmen, die nicht auf den Datenbank-User oder die eindeutige Identifikation der Session, die wiederum Rückschlüsse zugelassen hätte, beruhen. Anstelle der Session-ID wurde auf eine willkürliche, eineindeutige Sequenz zurückgegriffen, die ausschließlich einen Durchlauf eines Silentrecordings identifiziert. Aus den Produktionsdatenbanken wurden dann nur Transaktionsdaten inkl. eines Zeitstempels (bis auf Millisekunde genau) entnommen, d.h. User-Daten und Session-Identifikation wurden nicht extrahiert.

Die Zusammenführung der User-Interaktionen aus dem Silent-Recording und den zugehörigen Transaktionsdaten aus dem Datenbank-Log erfolgte dann über den Zeitstempel und dem Wissen über den Code, d.h. der Berücksichtigung der hochwahrscheinlichen Tatsache, dass nicht alle zeitgleich laufenden User Stories zur gleichen Millisekunde den gleichen schreibenden Zugriff auf die gleiche Tabelle haben.

AP 6.2 Prüfung der Machbarkeit der Einhaltung datenschutzrelevanter Vorgaben

Ausgehend von den in AP 6.1 konkret implementierten Maßnahmen und den erhobenen Daten wurden die Einhaltung der betriebs- und datenschutzrechtlichen Fragestellungen projektbegleitend geprüft und die Maßnahmen kontinuierlich weiterentwickelt, um auf diese Weise die Machbarkeit des Ansatzes nachzuweisen.

AP 6.3 Projektbegleitende Evaluation der Einhaltung datenschutzrelevanter Vorgaben

Ergänzend zu der datenschutzrechtlichen Prüfung bzw. dem Nachweis der Machbarkeit haben die PITSS und DBIS durch die in AP 6.1 beschriebenen technische und organisatorische Maßnahmen während der Implementierung die Einhaltung des Privacy-by-Design-Ansatzes laufend geprüft und sichergestellt.

Arbeitspaket AP7 - Entwicklung von Algorithmen für Process Mining und Process Analytics

AP 7.1: Entwicklung von Algorithmen für das ProcessTracking

Bei der Analyse der erfassten ActivityLogs wurde festgestellt, dass der Detaillierungsgrad an Informationen zu feingranular für die effektive Nutzung bestehender Process Mining Algorithmen ist. Entsprechend Arbeitspaket 3.2 (Entwicklung geeigneter Abstraktionsebenen und Sichten auf erfasste Prozesse) wurden daher verschiedene Verfahren entwickelt und evaluiert, mit denen sich wiederkehrende Aktivitäten auf kleinster Ebene automatisiert erkennen und bei Bedarf verdichten (d.h. geeignet abstrahieren) lassen. Diese Verfahren basieren auf Mustererkennungen in Graph Algorithmen und ermöglichen die Abstraktion von Logs ohne zusätzliche manuelle Interaktionen.

Zusätzlich ist sicherzustellen, dass neben der Verdichtung bzw. Abstraktion feingranularer Aktivitäten auch eine Anreicherung mit weiteren Metadaten (vgl. Ergebnisse zu der Beschreibung eines ActivityLogs in AP 3.2) möglich ist.

Außerdem wurde ein erster Entwurf für das Tracking von (objektzentrierten) Prozessen, wie sie in zahlreichen (Alt-)Software-Applikationen vorzufinden sind, in Form eines objektzentrierten Dashboards entwickelt (vgl. Publikation „A Dashboard-based Approach for Monitoring Object-Aware Processes“). Dieses wurde über den Projektzeitraum erweitert und auf Projektanforderungen angepasst.

AP 7.2: Entwicklung von Algorithmen für ein Conformance Checking

Im AP 7.2 wurde ein Ansatz entwickelt, um für datengetriebene, objektzentrierte Prozesse die Konformität (engl. *Conformance*) einzelner Objektinstanzen mit gegebenen Objektlebenszyklusprozessen zu überprüfen. Dabei ermöglicht der Ansatz, die inhärente Flexibilität im Verhalten von Prozessen zu berücksichtigen und dadurch mögliche und erlaubte Abweichungen vom idealen Prozessablauf zu erkennen (vgl. Publikation „Enabling Conformance Checking for Object Lifecycle Processes“). Es wurden zwei Granularitätsebenen für die Ausführung von Lebenszyklusprozessen und Konzepte für deren inhärente Ausführungsflexibilität eingeführt. Diese werden anschließend beim *Conformance Checking* einbezogen, um zwischen verschiedenen Ausführungsmöglichkeiten zu unterscheiden. Um dies zu ermöglichen, werden die Lebenszyklusprozesse in ausführungsäquivalente, formale Darstellungen – sog. Workflow-Netze – transformiert, die dann auf Conformance mittels traditioneller Conformance Checking Techniken überprüft werden. Die Prüfung der Konformität mit mehreren Netzen ermöglicht die Kategorisierung jeder Lebenszyklusausführung auf Grundlage des im Log erfassten Verhaltens.

Außerdem wird die flexible Ausführbarkeit von Objektlebenszyklusprozessen durch Konformitätskategorien berücksichtigt. Letztere ermöglichen es, zwischen tolerierten Abweichungen, die aufgrund der inhärenten Flexibilität möglich sind, und tatsächlichen Abweichungen (z. B. Ad-hoc-Änderungen), die spontan erforderlich wurden, zu unterscheiden. Zusätzlich ist der entwickelte Ansatz in der Lage, die Flexibilität bei Übergängen zwischen Zuständen (z. B. Änderungen in Formularreihenfolgen) und das Verhalten einzelner Zustände (z. B. Flexibilität beim Ausfüllen eines Formulars bzw. einer Maske) zu berücksichtigen. Konformitätskategorien geben Prozessverantwortlichen außerdem Hinweise darauf, welche Teile des Objektverhaltens (d.h. des Ausführungsverhaltens des zugehörigen Objektlebenszyklusprozesses) von besonderem Interesse für Verbesserungen sind.

AP 7.3: Entwicklung von Algorithmen für Compliance-Analysen von Instanzen datenobjekt- und benutzerzentrierter Prozesse

Das in AP 7.2 entwickelte Conformance Checking stellt in gewisser Weise schon die Überprüfung von Compliance Regeln dar, sofern diese im Verhalten des Objektes spezifiziert sind. Sollten weitere Compliance Analysen benötigt werden, d.h., Regeln die nicht im Prozessmodell spezifiziert sind, können die Workflownetze entsprechend erweitert werden. Dadurch wird auch ein Online-Monitoring von Compliance-Regeln (z.B. Einhaltung des 4-Augen-Prinzips bzgl. der Durchführung verschiedener Aktivitäten bzw. Aktionen) möglich, wodurch z.B. erkannt werden kann, welche Regeln erfüllt, noch nicht erfüllt oder dauerhaft verletzt sind.

AP 7.4: Implementierung und Testen der entwickelten Algorithmen

Die Algorithmen für AP 7.2 wurden vollständig implementiert und mit gängigen Verfahren evaluiert und getestet. Sie wurden von DBIS ferner auf der 16th International Conference on Research Challenges in Information Science präsentiert und publiziert (vgl. Publikation „Enabling Conformance Checking for Object Lifecycle Processes“).

Arbeitspaket AP8 - Integration der Algorithmen in Process Mining Tool

AP 8.1: Integration der Process Mining Algorithmen in ein Open-Source Process Mining Tool

Durch die Fähigkeit, den Eventstream in einem fest vordefinierten Format zu erstellen bzw. zu übergeben, können verschiedenste Prozess-Mining Algorithmen für die Identifikation, Entdeckung, Überprüfung und Verbesserung von Geschäftsprozessen eingesetzt werden.

Exemplarisch wurde im SoftProc-Projekt das Werkzeug pm4py (Process Mining for Python) evaluiert und für die Generierung von Prozessmodellen in verschiedener Granularität genutzt. Eine Übernahme dieser Bibliothek in die Gesamtkomponente wurde evaluiert und mit dem Owner (Fraunhofer Institute for Applied Information Technology (FIT)) besprochen. Das Lizenzmodell von pm4py sieht eine Open- sowie auch eine Closed-source Usage vor.

Für den Einsatz im Produkt SoftProc war zum Zeitpunkt der Veröffentlichung der Ergebnisse noch keine finale Lösung verhandelt.

AP 8.2: Konzeption und Implementierung der Visualisierungsfunktionen des Process Mining Tools

Aufgrund der Lizenzbeschränkungen von pm4py für kommerzielle Zwecke wurde ein eigenes Tool zur Analyse der Data-Activity-Log entwickelt welches die verschiedenen Cluster innerhalb des Altsystems abbildet.

Dabei wurde auf eine einfache Integration in das bereits vorhandene User-Interface der Produktpalette PITSS.CON (basierend auf dem Low-Code-Framework Oracle APEX) geachtet. Die Entscheidung fiel auf die OpenSource JavaScript Library Cytoscape (<https://js.cytoscape.org/>), die sich auf die Visualisierung von Graphen innerhalb einer WebApplikation spezialisiert hat. Im Backend wurden die Optionen einer relationalen, wie auch graphischen Datenhaltung über die Oracle Graph Database Engine offengehalten.

Parallel zu diesem Ansatz wurde gleichfalls eine klassische BPMN-modellierende Repräsentation eines Geschäftsprozesses über eine Schnittstellenfunktionalität (basierend auf Flows

for APEX <https://flowsforapex.org/>) enabled und somit für eine Darstellung im Produkt verfügbar gemacht.

Arbeitspaket AP9 - Erweiterung um Process Online Analytics

AP 9.1: Entwicklung und Implementierung von Online-Conformance-Checking Funktionen.

Das Modell zur Erhebung der geschäftsprozessrelevanten Informationen wurde von Anfang an so konzipiert, dass die Erfassung, Transport, Transformation und Interpretation automatisiert als Hintergrundprozesse ausgeführt werden können. Damit ist eine Echtzeit-Repräsentation der Daten z.B. als Dashboard möglich.

Die aufbereiteten Daten werden, wie schon im Arbeitspaket 4 beschrieben, als XES (eXtensible EventStream) zur Verfügung gestellt, so dass jedwelche Process Mining Werkzeuge auf dieser Basis Conformance Checking durchführen können.

Im Rahmen des Projektes PHILharmonic Flows wurde dieser Conformance Check umgesetzt. Dabei konnten die Prozess-Varianten, die über eine dedizierte Aufzeichnung erfasst wurde, als Mustervorlage der Prozesse interpretiert und anschließend mit den Aufzeichnungen aus den Silent-Recording verglichen werden.

Eine vielversprechende Erweiterung zeigt sich in der Möglichkeit, die Software-Variante (z.B. ein Release oder ein Patch) als Kriterium für den Conformance Check zu berücksichtigen. Gerade durch das Parsen der gesamten Applikation gelingt SoftProc eine eindeutige Zuordnung einer Prozess-Instance zu möglichen Versionen der eingesetzten Software. Damit könnten Rückschlüsse gezogen werden, ob eine Veränderung der Software auch zur Veränderung bzw. Verbesserung der Geschäftsprozesse (schnellerer Durchsatz, höhere Conformität, etc) geführt hat.

AP 9.2: Entwicklung und Implementierung ausgewählter Compliance Prüfalgorithmen, um Online Process Compliance Monitoring zu ermöglichen

Im Gegensatz zu den Conformance Checks wurde bei der Compliance Prüfung darauf verzichtet, die Daten für bereits vorhandene Systeme aufzubereiten. Die Erfassung der geschäftsprozessrelevanten Daten erlaubt die Berücksichtigung verschiedener Kenngrößen bei der Definition von Compliance Regeln. Diese sind vor allem Geschäftsobjekt-Entitäten mit allen bekannten Attributen, Benutzer-Interaktionen mit dem System und die Zeit. Mit diesen Informationen konnten Regeln definiert werden, wie z.B. „Alle Kundenaufträge mit dem Attribut „Auslandskunde“ sind spätestens 3 Tage nach Erfassung im Formular „Freigabe“ als genehmigt zu markieren.“

Die Aktion „markieren als genehmigt“ wird durch SoftProc als User-Interaktion erfasst. Damit verbunden ist eine Änderung an der Geschäftsprozess-Entität des Auftrages, die eindeutig der Aktion im Userinterface zugeordnet werden kann. Gleichzeitig speichert SoftProc für die eindeutige Kennung des Auftrages, so dass in der Historie der Aktionen zu dem Auftrag die Zeitdifferenz zur Erfassung, wie auch zu jeder anderen Aktion mit dem Auftrag ermittelt werden kann.

Die Konfiguration für Compliance Regeln ist bislang nur auf Basis von Datenbank-Tabellen und Prozeduren implementiert, die aber das wesentliche Backend für diese Funktion darstellen. Ein intuitives UserInterface, so dass ein prozesskundlicher Anwender neue Compliance Regeln definieren oder bestehende verändern kann, ist für die nächste Version eingeplant.

Arbeitspaket AP10 - Schnittstellenkonfiguration und Implementierung des Demonstrators

Um die Ergebnisse von SoftProc gerade in Kundenpräsentationen und für den Presales-Prozess visualisieren zu können, wurden eine Reihe von Schnittstellen und Demonstrationsumgebungen erstellt.

Für eine bessere Adaption und ein größeres Verständnis für den Funktionsumfang des Projektes, wurde sehr großer Wert darauf gelegt, die Funktionalitäten auf kundenspezifischen Sourcen präsentieren zu können.

Die Übergabe der kundenspezifischen Implementierung ist in der Regel keine große Herausforderung für den Kunden, da diese Informationen meist jetzt schon in externen Versionsverwaltungen oder zumindest im Filesystem vorhanden sind. Problematischer ist die Übergabe der Datenbank-Sourcen (Datenstruktur, Datenbanknahe Programmierung, etc) da hier der Kunde, gerade bei Inhouse-Lösungen, selten ein Konzept hat, diese Umgebungen einfach und schnell replizieren zu können. Daher wurde ein Verfahren entwickelt, welches sehr leichtgewichtig auf einer Installation des Kunden ausgeführt werden kann und als Ergebnis eine Schnittstellen-Datei zur Verfügung stellt, mit der alle notwendigen Informationen für die erste Statische Analyse der Applikation extrahiert werden können.

Diese ist die Basis für die im Arbeitspaket 5 beschriebene Ermittlung des Geschäftsprozess Entitäten.

Weitere Schnittstellen für die vollumfängliche Nutzung von SoftProc wurden erstellt, um die Datenbank-Log-Files sowie die Logfiles des Applikationsservers zu integrieren. Diese basieren auf Standards, die vom Hersteller der Infrastruktur vorgegeben werden.

Bei einer Nutzung von SoftProc in einem Kundenumfeld werden diese Schnittstellen in die Infrastruktur des Kunden integriert. Dazu gehören die Definition der Zugangspfade, ggfls. zu vergebende Berechtigungen, sowie die Anpassung der Backend-Prozesse um Extraktion, Transformation und Interpretation der Daten zu automatisieren.

SoftProc als Gesamtheit wurde als Software-as-a-Service konzipiert und kann sowohl als Cloud-Lösung als auch als on-prem-Lösung zum Einsatz kommen. Innerhalb der Cloud-Variante wurde ein Beispiel-Szenario vorbereitet, um die Funktionsweise des Produktes zu demonstrieren.

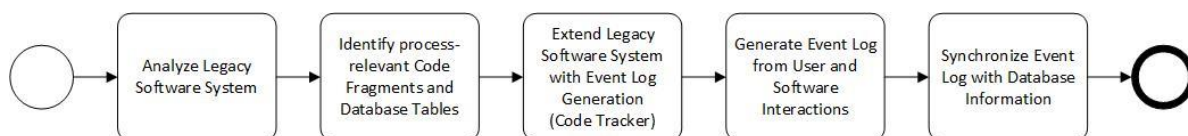


Abbildung 5: Im Rahmen des Vorhabens entwickelter Ablauf

Arbeitspaket AP11 – Projektbegleitende Evaluation

Im Arbeitspaket AP 11 wurden seitens der PITSS parallel zu den Entwicklungsarbeiten umfassende Funktions- und Leistungstests der entwickelten Softwarekomponenten unter Realbedingungen durchgeführt. Daran anknüpfend wurde eine Szenario-Analyse zur Optimierung der Lösung erarbeitet.

Nach der Absage des assoziierten Verbundkooperationspartner sensis lag dem Vorhaben zunächst kein konkreter Kundenbezug vor. Um die Arbeitspakete dennoch in dem beantragten Umfang durchführen zu können, wurde eine typische, praxisnahe ERP-Applikation als Referenzinstallation aufgesetzt.

Der daraus resultierende Vorteil dieses Ansatzes ergab sich in der Synergie, dass innerhalb des SoSe-Semesters die Student*innen beim Partner DBIS sowohl eine Modellierung des Prozesses mit den verschiedenen oben genannten Verfahren BPMN und PHILharmonic Flows durchführten, als auch die modellierten Prozesse am System selbst abarbeiten konnten.

Die weiteren Arbeiten des AP 11 beinhalteten:

- Aufstellung von praxisnahen Testszenarien
- Durchführung von projektinternen Funktionstests auf Basis einer ERP-Applikation, die sich bei einer Reihe von PITSS-Kunden im Einsatz befindet und dessen Sourcen für die PITSS GmbH verfügbar sind.

Arbeitspaket AP12 - Öffentlichkeitsarbeit und Transfer

Im Rahmen des Arbeitspakets AP 12 wurden von den Entwicklungspartnern PITSS und DBIS die Forschungsergebnisse ausgewertet, geeignet aufbereitet und veröffentlicht.

Das DBIS hat zudem die Projekt-Ergebnisse in diversen wissenschaftlichen Arbeiten veröffentlicht. Darüber hinaus wurden sie auf Konferenzen und Workshops vorgestellt. Im Detail entstanden im Verlauf des Projektes eine Reihe von projektbezogenen wissenschaftlichen Veröffentlichungen. (Vgl. Kapitel 7)

Während der Fokus des DBIS auf der wissenschaftlichen Verbreitung der Forschungsergebnisse lag, hat sich PITSS auf die allgemeinverständliche Kommunikation der Ergebnisse fokussiert.

Unter der URL bzw. Landingpage <https://pitss.com/de/retrograde-prozessanalyse-leicht-gemacht/> wurde im Projektverlauf auf die Projektergebnisse hingewiesen.

3. Wichtigsten Positionen des zahlenmäßigen Nachweises

3.1. PITSS GmbH

Die geplanten Mittel wurden entsprechend der ursprünglichen Planung seitens der PITSS GmbH weitestgehend eingehalten. Entsprechend wurden Softwareentwickler auf das Projekt eingesetzt.

Die Reisemittel wurden aufgrund der Corona-Pandemie nicht in der geplanten Form aufgewendet.

3.2. Universität Ulm

Im SoftProc-Projekt haben seitens des Instituts für Datenbanken und Informationssysteme (DBIS) der Universität Ulm – wie geplant – der Projektleiter Prof. Dr. Manfred Reichert sowie die wissenschaftlichen Mitarbeiter*innen M.Sc. Marius Breitmayer und M.Sc. Lisa Arnold aktiv mitgewirkt. Dies unterstreichen auch die aus dem Projekt hervorgegangenen wissenschaftlichen Veröffentlichungen (s. Publikationsliste).

Hinzu kommen Studierende, die im Rahmen ihrer Projekte und Abschlussarbeiten (Bachelor, Master) an der Bearbeitung bestimmter Teilaspekte im Projekt aktiv beteiligt waren.

Die tatsächlich entstandenen Ist-Ausgaben der Universität Ulm im Projekt lagen geringfügig unterhalb der beantragten Mittel. Wie dargestellt, konnte der Arbeitsplan dennoch vollumfänglich eingehalten und umgesetzt werden.

4. Notwendigkeit und Angemessenheit der Projektarbeiten

4.1. PITSS GmbH

Ausgehend von der Zusammenarbeit in dem Projekt war es der PITSS möglich, die vorhandenen Kompetenzen im Bereich der Softwaremodernisierung noch einmal erheblich auszubauen.

Auch ist im Verlauf des Projektes noch einmal deutlich geworden, dass die Entwicklung einer Software sowie von professionellen, softwaregestützten Dienstleistungen nicht auf die leichte Schulter genommen werden kann und von einem KMU wie PITSS, in der Form und Zeit, hätte ohne die Unterstützung des Projektpartners DBIS nicht durchgeführt werden können. Die Entwicklung und prototypische Implementierung von SoftProc im Kontext komplexer IT-Umgebungen erforderten einen ganzheitlichen Ansatz beider am Vorhaben beteiligten Partner sowie die sich daraus ergebene Symbiose aus (wissenschaftlicher) Theorie und unternehmerischer Praxis.

Für die PITSS eröffnet die im Projekt erforschte und umgesetzte Herangehensweise die Tür für den Einsatz von Machine Learning und KI-Verfahren im Spannungsfeld von zum einen existierenden Datenströmen und Log-Informationen und zum anderen den dazu notwendigen Modellen und Implementierungen.

Mit Blick auf die wirtschaftliche Verwertung sowie Pflege und Weiterentwicklung der Projektergebnisse wurden bei der PITSS bestehende Entwicklerstellen verstetigt sowie weitere Stellen neu geschaffen.

4.2. Universität Ulm

Im Projekt hat DBIS bzw. die Universität Ulm neuartige Verfahren zur Online-Analyse objektzentrierter Prozesse entwickelt, die ohne die Förderung nicht hätten entwickelt werden können. Eine wesentliche Unsicherheit vor dem Projekt bestand darin, ob die zu entwickelnden Verfahren mit den in Unternehmen vorliegenden Prozessdaten sinnvolle und qualitativ hochwertige Aussagen treffen können, die für den Einsatz in realen Umgebungen nutzbar sind. Insbesondere der Einsatz solcher Verfahren in komplexen und verzweigungsreichen Softwareumgebungen war nicht abzuschätzen. Daher wäre ohne Förderung das wissenschaftlich-technische Risiko zu hoch gewesen, um in diesem Bereich die angedachten Lösungen zu entwickeln.

Der Austausch mit dem KMU bzw. Industriepartner im Projekt hat weiter dazu beigetragen, ein tiefes Verständnis für die in der wirtschaftlichen Praxis bzw. in Unternehmen vorliegenden Daten zu erhalten. Ohne den Einbezug der PITSS hätte die Universität Ulm in diesem Bereich keine wesentliche wissenschaftliche Forschung betreiben bzw. diese nicht angemessen evaluieren können.

5. Voraussichtlicher Nutzen, insbesondere Verwertbarkeit

5.1. PITSS GmbH

Das Projekt führte bei der PITSS insgesamt zu einer deutlichen Verschärfung des Profils der begleitenden Dienstleistungen im Bereich der Softwaremodernisierung: Die entwickelten Module / Schnittstellen und Lösungen werden für das nächsten Release für Anfang 2024 adaptiert und werden zum Einsatz kommen.

Den Projektergebnissen werden auch weiterhin hinsichtlich der Verwertung große Erfolgsaussichten zugeschrieben, da in der aktuellen Situation der Software-Entwicklung, gerade im Bereich der Wartung- und Pflege von Legacy-Applikationen noch immer ein Investitionsstau zu erkennen ist: zahlreiche Legacy-Applikationen, auf die dieses Projektvorhaben abzielen, basieren auf einer Technologie, die bei der Erweiterung von Geschäftsfeldern, den Kunden unüberwindbare Hindernisse in den Weg stellen. Lösungsräume, die durch Cloud-Lösungen oder nativen Webapplikationen möglich wären, verschließen sich für diese Legacy-Applikationen und sehen als einzigen Ausweg komplette Neuentwicklungen der Anwendung.

Durch die Entwicklungsarbeiten und die daraus gewonnenen Erkenntnisse ist für das Unternehmen nun eine deutlich breitere Marktbearbeitung möglich: einige Bereiche bzw. Branchen, die vorher aufgrund veralteter komplexer Softwareumgebungen nur sehr begrenzte Möglichkeiten zur Modernisierung ihrer Systeme hatten, erhalten diese Möglichkeit nun über Projektergebnisse, da mit der neuen Lösung erheblich weniger technisches Process Mining und Beratungsaufwand notwendig ist.

Gleichwohl sind diese allerdings auch mit der neuen Lösung mit Investitionen und Risiken verbunden, so dass in der aktuellen – mitunter schwierigen – wirtschaftlichen Gesamtsituation konkrete Projektumsetzungen aktuell eher gescheut werden.

Das Projektvorhaben eröffnet der PITSS in einem ersten Schritt, mit der Identifikation der Geschäftsprozesse die Möglichkeit, gezielt und planungssicher die Anwendung unter Berücksichtigung eben dieser Business Prozesse zu optimieren und durch die gleichzeitige Verbindung mit den zugehörigen Sourcen und Datenmodellen der Anwendung, diese gezielt in eine neue Technologie zu überführen.

Aktuelle Ansätze in den Konkurrenzlösungen sehen jeweils nur einen Teil dieser Aufgabenkette beim Kunden im Fokus, so dass weiterhin hoher Transformationsaufwand für den Kunden bestehen bleiben würde.

Innerhalb der PITSS wurden im Kontext des SoftProc-Projektes neue Bereiche geschaffen die ca. 2 Arbeitsplätze zusätzlich erzeugt haben.

Mit Blick auf die wirtschaftliche wie auch wissenschaftlich-technische Verwertung lassen sich die wesentlichen Projektergebnisse auf Seiten der PITSS wie folgt darstellen:

Analyse über Datenmodell

Für das Projektvorhaben war es notwendig, den Änderungen am Datenmodell, die durch die Benutzerinteraktionen ausgelöst wurden, eine semantische Bedeutung im Umfeld der zu identifizierenden Geschäftsprozesse zuzuordnen. Dazu war es notwendig, innerhalb des zunächst unbekanntes Datenmodell des Kunden automatisiert eine Semantik zu erkennen.

Innerhalb des Projektes konnte ein Verfahren entwickelt werden, mit dem über dem ERM (Entity-Relationship-Model) ein weiteres Modell retrograd erstellt wurde, welches zum Einen zusätzliche Relationen zwischen den Entitäten innerhalb der Implementierung und zum Anderen eine Klassifikation der Entitäten basierend auf der Anzahl der DML (Data-Manupulation-Language) Statements erkennen läßt. Dadurch war ein Clustering des Datenmodells und die Benennung der Cluster-Repräsentanten ohne Applikationswissen möglich. Basierend auf diesem erweiterten Modell konnte automatisch erkannt werden, welche Spalten für die Identifikation der Geschäftsobjekte zwingend notwendig sind.

Nutzbarkeit von Applikation und Database-Server Logfiles

Es konnte eine Alternative zum invasiven Tracken der Benutzerinteraktionen identifiziert werden. Mit den Logfiles des Applikation Server und der Datenbank konnte nahezu der gleiche Detailgrad an Informationen ermittelt werden, wie mit einer automatisierten Änderung der Software zur Integration der Tracking-Informationen.

Dieses Verfahren kann additiv genutzt werden, um z.B. weitere Informationen aus einer Produktiv-Umgebung zu erheben, ohne Änderungen (Change) an der existierenden Codebase.

Erweiterung von Eventstreams für Aktionen, die direkt aus dem Framework erfolgen

Das im Fokus stehende Entwicklungsframework „Oracle Forms“ ermöglicht dem Entwickler ohne Implementierung von PL/SQL DML-Statements innerhalb der Applikation zu nutzen. Um diese Änderungen gleichfalls analysieren zu können, wurde ein Verfahren entwickelt, um auch in diesen Fällen die zugehörige User-Interaktion und die Änderungen an den Applikationsdaten zu zuordnen. Damit konnte eine vollständige Abdeckung aller DML-Statements innerhalb der Applikation erreicht werden.

Fehlerfreie Zuordnung von User-Interaktionen zu Modifikationen im Datenmodell

Durch die Verbindung einer Session-ID zwischen dem Applikations- und Datenbanklogfile, sowie das Festhalten der im User Interface aktuell genutzten Geschäftsobjekte konnte sichergestellt werden, dass fehlerfrei eine User-Interaktion auch den Änderungen in den Applikationsdaten zugeordnet werden konnte.

Verdichtung von feingranularen Mikro-Events zu Business Processen.

Durch das Verfahren der Eventlog-Abstraction konnten trotz der Vielzahl der feingranularen Events, die in einem einzelnen Prozess auftauchen können, die wesentlichen Schritte innerhalb eines Prozesses ermittelt und visualisiert werden. Diese beziehen sich auf Navigation zwischen verschiedenen Formularen, Teile des User-Interfaces oder Auslöser, wie z.B. Befehlsflächen.

Diese Abstraktion gelingt rein auf Basis der Software-Struktur und der Erkennung der Implementierung ohne notwendiges Applikationswissen.

Transformation der Implementierung in eine natürliche Prozessbeschreibung

Da auch die Implementierung des reinen User-Interface, wie z.B. die Beschriftung der Befehlsflächen, die Überschriften an Tabellen und Spalten oder auch Fenstertitel Bestandteil der automatisierten Software Analyse waren, konnten diese Beschriftungen für eine anschließende automatische Beschreibung des Prozesses als natürlich-sprachliche Beschreibung benutzt werden.

So kann die individuelle „Sprache“ der Applikation aufgegriffen und die Kommunikation mit den Anwendern deutlich vereinfacht werden.

Alle angesprochenen Ergebnisse und Erkenntnisse werden in das kommende Release des Produktes PITSS.CON einfließen und stehen somit der Kundenbasis als update oder auch für Neukunden als eigenständiges Produkt zur Verfügung.

5.2. Universität Ulm

Die Universität Ulm sieht weiterhin große Potenziale im Bereich der Prozessanalyse, insbesondere der Analyse objektzentrierter Prozesse, und verwertet – in Abstimmung mit der PITSS sowie unter Berücksichtigung unternehmerischer Interessen – ausgewählte Forschungsergebnisse in Form von Präsentationen und Publikationen auf internationalen Fachkonferenzen & Workshops sowie Veröffentlichungen in renommierten Zeitschriften. Nicht zuletzt durch das SoftProc-Projekt konnte DBIS bzw. die Universität Ulm ihre international herausragende Rolle in den Bereichen Business Process Management und Process Mining stärken.

Im Rahmen des Projektes wurden insgesamt 12 Publikationen veröffentlicht. Die Publikationen beschäftigen sich mit den Forschungsbereichen Process & Data Mining. Sie beschreiben originelle Konzepte, Implementierungen und Evaluationen, die im SoftProc Projekt durchgeführt wurden.

Darüber hinaus hat das DBIS ausgewählte Projektergebnisse in der forschungsnahen Lehre (u.a. in den Lehrmodulen zu „Business Process Intelligence“, „Business Process Management“, „Business Process Modeling Lab“, „DBIS-Lab“ und „Process & Data Mining“) integriert. Ebenso wurden ausgewählte Ergebnisse in eingeladenen Präsentationen des DBIS-Institutsleiters Prof. Reichert auf Fachkonferenzen und in Forschungskolloquien präsentiert. Prof. Reichert war zudem Programmverantwortlicher der 19. Internationalen Fachtagung für Business Process Management (BPM 2021) im September 2021 in Rom, was die internationale Sichtbarkeit des DBIS und seiner Führungsrolle untermauert.

Im Rahmen des Fördervorhabens sind insgesamt 15 Bachelor- und Masterarbeiten zu den Themen Process Mining, Process Analytics, Online Conformance / Compliance Checking sowie Prozessqualität betreut worden. Das Projekt wurde folglich eng von Studierenden begleitet und unterstützt mit z.T. hervorragenden Ergebnissen.

Eine wirtschaftliche Verwertung seitens der Universität Ulm ist nicht vorgesehen.

6. Fortschritte in Wissenschaft und Technik während der Projektlaufzeit

Während der Durchführung des Vorhabens wurden folgende neue Arbeiten – nach projektrelevante Themenfeldern kategorisiert – dem ZE bekannt:

- **Object-centric Process Mining**

Objektzentriertes Prozess-Mining ist ein Ansatz zur Modellierung und Analyse von Geschäftsprozessen, der den Fokus auf individuelle Objekte legt, die durch den Prozess fließen. Im traditionellen Prozess-Mining liegt der Schwerpunkt oft auf der Abfolge von Aktivitäten in einem Prozess. Objektzentriertes Prozess-Mining hingegen betrachtet, wie einzelne Objekte oder Datensätze (wie Bestellungen, Kundenanfragen oder Patientenakten) durch einen umfangreichen Geschäftsprozess wandern.

Der Projektpartner DBIS hat sich dank des SoftProc Projektes frühzeitig in diesem neuen Process Mining Feld eingebracht. Durch die Vorarbeiten in Form von PHILharmonicFlows ist DBIS nun nicht nur in der Lage die Analyse solcher Prozesse zu unterstützen, sondern den gesamten BPM-Lebenszyklus (das bedeutet die Phasen Discovery, Analysis, Neugestaltung, Implementation, sowie Überwachung) in Form eines Gesamtframeworks abzudecken.

7. Erfolgte und geplante Veröffentlichungen

Im Verlauf des Projektes entstanden folgende projektbezogenen wissenschaftlichen Veröffentlichungen:

1. Breitmayer, Marius and Arnold, Lisa and Reichert, Manfred (2021) A Dashboard-based Approach for Monitoring Object-Aware Processes. In: 13th Central European Workshop on Services and their Composition, Bamberg.
2. Arnold, Lisa and Breitmayer, Marius and Reichert, Manfred (2021) A One-Dimensional Kalman Filter for Real-Time Progress Prediction in Object Lifecycle Processes. In: IEEE Int'l Workshop on Intelligent Digital Architecture, Methods, and Services for Industry 4.0 and Society 5.0 (IDAMS 2021), EDOC 2021 Workshop, Gold Coast, Australia, 25 - 26 October 2021, IEEE Computer Society Press.
3. Arnold, Lisa and Breitmayer, Marius and Reichert, Manfred (2021) Towards Real-Time Progress Determination of Object-Aware Business Processes. In: 13th Central European Workshop on Services and their Composition, Bamberg.
4. Breitmayer, Marius and Arnold, Lisa and Reichert, Manfred (2022) Towards Retrograde Process Analysis in Running Legacy Applications. In: 14th Central European Workshop on Services and their Composition, February 24-25, Hannover.
5. Breitmayer, Marius and Arnold, Lisa and Reichert, Manfred (2022) Enabling Conformance Checking for Object Lifecycle Processes. In: 16th International Conference on Research Challenges in Information Science (RCIS 22), Barcelona, Spain, 17-20 May 2022, LNBIP, Springer, pp. 124-131.
6. Arnold, Lisa and Breitmayer, Marius and Reichert, Manfred (2022) Progress Determination of a BPM Tool with Ad-Hoc Changes: An Empirical Study. In: 16th International Conference on Research Challenges in Information Science (RCIS 22), Barcelona, Spain, 17-20 May 2022, LNCS, Springer, pp. 107-123.
7. Breitmayer, Marius and Arnold, Lisa and Reichert, Manfred (2022) Data-Driven Evolution of Activity Forms in Object- and Process-Aware Information Systems. In: 27th International Conference on Cooperative Information Systems (CoopIS 2022), Bozen - Bolzano, 4 -7 October 2022, LNCS 13591, Springer, pp. 186-204.
8. Breitmayer, Marius and Arnold, Lisa and La Rocca, Stephan and Reichert, Manfred (2022) Deriving Event Logs from Legacy Software Systems. In: Workshop at 4th International Conference on Process Mining (ICPM 2022), Bozen - Bolzano, 23 – 28 October 2022, ICPM 2022 Workshops, Springer.
9. Breitmayer, Marius and Arnold, Lisa and La Rocca, Stephan and Reichert, Manfred (2023) Providing Hierarchy-Based Process Models from Legacy Software Systems. In: 15th Central European Workshop on Services and their Composition (ZEUS 2023), Hannover

10. Arnold, Lisa and Breitmayer, Marius and Reichert, Manfred (2023) Towards Progress Determination in Dynamically Evolving Large Process Structures. In: 15th Central European Workshop on Services and their Composition (ZEUS 2023), Hannover
11. Arnold, Lisa, Marius Breitmayer, and Manfred Reichert (2023). Monitoring Object-Centric Business Processes: An Empirical Study. International Conference on Research Challenges in Information Science. (RCIS 23), Korfu, Springer Nature Switzerland.
12. Breitmayer, Marius and Arnold, Lisa and Reichert, Manfred (2023) Data-Driven Customization of Object Lifecycle Processes. In: 25th IEEE International Conference on Business Informatics (CBI 2023), Prague

Neben den aufgelisteten Veröffentlichungen befinden sich derzeit noch weitere Artikel in der Begutachtung:

1. Breitmayer, M., Arnold, L., Winter, M. and Reichert M. Exploring User Editing Preferences in Electronic Forms: An Empirical Study (In Revision)
2. Arnold, L., Breitmayer, M and Reichert, M. Determining the Progress of Business Objects

Kurzbericht zum Vorhaben

SoftProc Entwicklung von Verfahren zur retrograden Prozessanalyse aus einer laufenden Applikation heraus

Zuwendungsempfänger: PITSS GmbH Universität Ulm	Förderkennzeichen: 01IS20027A-B
---	---

Laufzeit des Vorhabens: **01.04.2021 bis 31.03.2023**

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS20027A-B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim der Autorin/beim Autor.

1.1. Ursprüngliche Aufgabenstellung

Die Implementierung von Softwareanwendungen adressiert die Bedürfnisse von Benutzern, Anwendungsfällen und Geschäftsprozessen. Gängige Software-Systeme sind jedoch mehrheitlich nicht darauf ausgelegt, prozessbezogene Daten hoher Qualität bereitzustellen, um mit modernen Process Mining Werkzeugen umfassende Prozessanalysen und -visualisierungen vorzunehmen. Relevante Fragestellungen sind z.B., wie der von der Software implementierte Prozess aussieht (Process Discovery) und wie stark dieser vom definierten Soll-Prozess abweicht (Conformance Checking).

Da in den meisten Fällen die Individualsoftwarelösung nicht zwingend eine Workflow- bzw. Prozess-Engine im Hintergrund nutzt, die mit wenigen Erweiterungen hilfreiche Audit-Daten zu liefern vermag, bleibt vielfach nur der Blick auf die Veränderungen im Datenmodell. Diese sind allerdings mit viel Zeitverlust durch Interviews, Analysen und Vermutungen verbunden.

1.2. Zielstellung

Ziel des Vorhabens war die Entwicklung softwaregestützter Verfahren, die Anwender bei der Abarbeitung eines Prozesses mit der Software „beobachten“ und diese Informationen feingranular in einen Event-stream bzw. ein Event Log schreiben. Mit der resultierenden Online-Prozessanalyse werden aus einer laufenden Applikation dedizierte Prozesse retrograd ermittelt und die Verwendung der Anwendung über unterschiedliche Anwender, Standorte oder Zeitpunkte dokumentiert.

1.3. Wissenschaftlicher und technischer Startpunkt

Die gängigen Ansätze, um zu Modellen für Geschäftsprozesse zu gelangen, haben gemeinsam, dass die zu analysierenden Software-Systeme einen möglichst akkuraten Eventstream an ein Process Mining Tool übergeben wollen. In 2016 hat die IEEE das Standardformat XES (Extensible Event Stream) für Eventstreams definiert. Dieser Standard wird von gängigen Process Mining Werkzeugen unterstützt, er erlaubt u.a. eine freie Erweiterbarkeit mittels eigener Extensions, um Spezifika des Unternehmens bzw. seiner Geschäftsprozesse abzubilden, ebenso wie Spezifika der Software-Systeme und die Eventstreams an das Process Mining Tool zu übergeben. Für die großen Standardsysteme mit konstanten Datenmodellen (z.B. SAP, Microsoft Dynamics, Oracle E-Business Suite) gibt es bereits vorkonfigurierte Exportstrecken, die z.B. für einen klassischen Beschaffungsprozess alle notwendigen Eventstream-Informationen aus der Applikation extrahieren. Jedoch gibt es keine simplen Exportstrecken für Standard-Anwendungen oder Branchenlösungen, geschweige denn für Individualsoftware.

Da in den meisten Fällen die Individualsoftwarelösung nicht zwingend eine Workflow- bzw. Prozess-Engine im Hintergrund nutzt, die mit wenigen Erweiterungen hilfreiche Audit-Daten zu liefern vermag, bleibt vielfach nur der Blick auf die Veränderungen im Datenmodell. Diese sind allerdings, wie bereits oben erörtert, mit viel Zeitverlust durch Interviews, Analysen und Vermutungen verbunden.

1.4. Wesentliche Ergebnisse

Im Rahmen der Forschungs- und Entwicklungsarbeiten wurde sichtbar, dass die größte Innovation innerhalb der Arbeitspakete in der Identifikation der Prozesse auf Basis von User-Interaktionen, Identifikation des "realen" Datenmodells und dem DataActivityLog liegt. Sofern daraus erst eine Prozessbeschreibung mit BPMN oder PHILharmonicFlows erstellt werden kann, waren die weiteren Innovationen im Bereich Konformität und Compliance nur noch geringfügig.

Aus diesem Grund lag im weiteren Projektverlauf der Fokus deutlich stärker auf einer hohen Qualität bei der semiautomatischen Erkennung der in der Anwendung implementierten Business-Prozesse.

1.4.1. PITSS GmbH

Die wesentlichen Ergebnisse auf Seiten der PITSS lassen sich wie folgt benennen:

- Entwicklung von Verfahren zur Analyse über das bestehende Datenmodell: innerhalb des zunächst unbekanntes Datenmodell kann automatisiert eine Semantik erkannt werden
- Entwicklung einer Alternative zum invasiven Tracken der Benutzerinteraktionen durch Nutzung von Logfiles von Applikation- und Database-Server
- Erweiterung von Eventstreams für Aktionen, die direkt aus dem Framework erfolgen ohne Implementierung von PL/SQL DML-Statements
- Fehlerfreie Zuordnung von User-Interaktionen zu Modifikationen im Datenmodell
- Verdichtung von feingranularen Mikro-Events zu Business-Prozessen.
- Transformation der Implementierung in eine natürliche Prozessbeschreibung

Alle angesprochenen Ergebnisse und Erkenntnisse werden in das kommende Release des Produktes PITSS.CON einfließen und stehen somit der Kundenbasis als Update oder auch für Neukunden als eigenständiges Produkt zur Verfügung.

1.4.2. Universität Ulm

Das DBIS hat in dem Vorhaben die PITSS mit neuesten Erkenntnissen aus der Wissenschaft und Forschung unterstützt und ihre entsprechende Expertise im Bereich Business Process Management in das Projekt eingebracht. Sie hat ferner die Algorithmen und Verfahren zur Feststellung der Konformität und Compliance implementierter Prozesse mit einem gegebenen Soll-Prozessmodell bzw. mit bestehenden Compliance-Regeln konzipiert und evaluiert.

Ferner hat DBIS die besonders originellen Ergebnisse seiner im SoftProc-Projekt geleisteten Forschungsarbeiten hochkarätig publiziert sowie auf 12 internationalen Konferenzen und Workshops präsentiert.