

EIS WS1516 - Meilenstein 2

Verteiltes Training einer automatisierten Dokumentenattributierung

Tim Howe

TAARs - Verteilte Group- Middleware zum Training einer automatisierten Attributierung von Rechnungen

Inhaltsverzeichnis

1 Zielhierarchie	3
1.1 Strategisch	3
1.2 Taktisch	3
1.3 Operational	3
2 Marktrecherche	4
2.1 Codia DMS	4
2.2 InPunkto	4
2.3 Übersicht	4
3 Domänenrecherche	5
3.1 Allgemeine Verarbeitungprozess	5
3.2 Der Begriff der Attributierung	5
3.3 Strukturierungsgrad	6
3.4 Anwendungsfelder	7
3.5 Nutzungsmerkmale	7
4 Alleinstellungsmerkmale	8
5 Methodischer Rahmen	9
6 Kommunikationsmodell	10
7 Risiken	11
7.1 Fachlich	11
7.2 Technisch	11
7.3 Marktzugang	11
8 Proof of Concepts	12
9 Architekturdiagramm	13
9.1 Übersicht	13
9.2 Verwaltungsdienst	13
9.3 Clients	13
9.3.1 Verwaltung	13
9.3.2 Fachclient	13
9.3.3 Steuerung	13
9.4 Regel-Engine	13
10 Projektplan	14
11 Projektbegründungen	15
11.1 Implementierung	15
11.2 Objektbereich	15
11.3 Nutzermodelle	15

Kapitel 1

Systembeschreibung

Ein allgemeine Beschreibung der wichtigsten Eigenschaften des TAARs , detaillierte Beschreibungen folgen in den zugehörigen Kapiteln

1.1 funktionale Komponenten

1.2 Akteure

1.3 •

Kapitel 2

Zielhierarchie

2.1 Strategisch

1. Anwendungskontext:
Es soll ein Anwendungskontext mit möglichst hoher wirtschaftlicher Relevanz gefunden werden
2. Technologisch:
Es sollen möglichst viele im beruflichen Kontext relevanten Erfahrungen... , siehe ??
3. Objektbereich:
Es muss ein Komplexitätsgrad erreicht werden der fachlich relevant und technologisch, im Rahmen des Projekts, beherrschbar ist
4. Nutzung:
Anwender sollen vom Ballast repetitiver Aufgaben befreit werden
- 5.

2.2 Taktisch

1. Objektbereich:
Es muss eine Analyse und Bewertung der in der Anwendungsdomäne genutzten Objekte, dh. dokumentenklassen, durchgeführt werden
2. Nutzung:
Es muss Automatisierungspotential identifiziert werden
3. ...
4. ...

2.3 Operational

1. Nutzung:
 - (a) deskriptive Aufgabenanalyse
 - (b) Automatisierungspotential identifizieren
 - (c) präskriptive Aufgabenanalyse, inkl P2
- 2.

Kapitel 3

Marktrecherche

Der Markt wird im allgemeinen von Entwicklern einzelner Komponenten und Systemhäusern bestimmt die Fremdkomponenten ggf mit Eigenentwicklungen kombinieren und so individuelle Lösungspakete schnüren.

Die Komponenten lassen sich grob in die folgenden Bereiche einteilen:

- OCR: Scannen, Klassifizieren
- Workflow
- betriebliche Anwendungssysteme: Buchhaltung, Archivierung

Zudem sind sogenannte OEM Versionen glossar durchaus gängige Praxis wodurch eine Einsicht erschwert wird. Um dennoch einen Eindruck in die Marktsituation zu gewinnen hilft eine Betrachtung der og Bereiche in der funktional übergeordneten Ebene der DMS -Systeme und deren Eigenschaften. Es werden Ansätze zweier Anbieter exemplarisch beschrieben und eine Einordnung zwischen diesen Lösungen und TAARs mittels einer Featurematrix ermöglicht.

3.1 Codia DMS

Codia DMS bietet auf Basis des d3.ecm von d.velop spezialisierte Lösungen im eGovernment Umfeld für öffentliche Verwaltung und Hochschulen mit den Themen Scannen Klassifizierung, Rechnungs- und Eingangspostverarbeitung, eAkte und Archivierung.

3.2 InPunkto

InPunkto spezialisiert auf Dokumenten Dienstleistungen im SAP Umfeld mit den Themen Automatische Erfassung & Verarbeitung, Workflow, eAkte und Archivierung.

3.3 Übersicht

Thema	Codia	InPunkto	TAARs
Automatisierte Klassifizierung	J	J	N
Attributierung semantisch	J	kA	N
fachlich	J	J	J
Automatisierte Attributierung	N	N	J
Steuerung	lastabhängige Aufgabenverteilung über Workflowsystem	kA	Priorisierung
(Rechnungs) Workflow	d3ecm	SAP Workflow	freie Wahl
Export	d3.ecm	SAP	Rohexport als xml

Kapitel 4

Domänenrecherche

Im folgenden werden die wichtigsten Themen der Anwendungsdomäne beschrieben und einer ersten Bewertung hinsichtlich projektrelevanter Eigenschaften unterzogen.

4.1 Allgemeine Verarbeitungprozess

Beispielhaft die wichtigsten funktionalen Komponenten des Dokumentenverarbeitungsprozesses beschrieben, wichtig bleibt anzumerken das in konkreten Implementierungen die Funktionalitäten verschwimmen und keine klare Trennung wie hier beschrieben vorherrscht(?).

1. Extraktion:
Manuelles oder mittels OCR automatisiertes auslesen von Informationen aus einer Dokumentendatei oder einer zugehörigen Bitmapdatei.

(a) optional Nacherfassung:
Kontrolle der OCR Ergebnisse und ggf Korrektur bei unzureichender Extraktionsqualität
2. Klassifizierung:
Beschreibt die Einordnung in eine klar abgegrenzte Menge von Dokumenttypen wie zb. Formulare, Rechnungen, Lieferscheine, Bewerbungen.
3. Attributierung:
Beschreibt den Prozess der Zuordnung von organisations- oder fachspezifischen Attributen zu einem Dokument zur weiteren Verarbeitung innerhalb der Organisation. Die können zum Beispiel Buchungskonten, Kostenstellen, Projektnummern oder Ansprechpartner sein.
4. Export & weitere Verarbeitung:
Übergabe der klassifizierten und attributierten Dokumente an einem betriebliches Anwendungssystem zur Buchung, Kontierung oder Archivierung.

4.2 Der Begriff der Attributierung

In der Praxis wird der Begriff der Attributierung für zweierlei Aspekte verwendet:

1. semantische(?) Attributierung:
Extrahierung und Zuordnung von Metainformationen zu einem Dokument welche sich auf das Dokument, bzw die Dokumentendatei als Repräsentation des Dokuments, selbst beziehen. Dies können zb Schlagworte, Speicherort ... sein.
2. fachliche Attributierung:
Die fachliche Attributierung ist ein Teil des Arbeitsprozesses bei welchem dem Dokument Attribute zugeordnet werden die Diese Attribute können zb kaufmännischer, steuerlicher oder juritischer Natur sein.

Für dieses Projekt wird der Begriff im Sinne der fachlichen Attributierung verwendet.

4.3 Strukturierungsgrad

Der Strukturierungsgrad eines Dokuments wird beschrieben durch das Ausmaß an Sicherheit mit der ein Wert eines Dokumentenattributs an einer Position im Dokument auftritt und welcher Wertebereich in diesem abgebildet wird.

1. unstrukturiert: gar keine bis geringe Positionssicherheit mit überwiegend undefiniertem Wertebereich, zb. Bewerbungsschreiben
2. semi-strukturiert: gute Positionssicherheit mit überwiegend definiertem Wertebereich, zb. Rechnungen, Lieferscheine
3. strukturiert: absolute Positionssicherheit mit klar definiertem Wertebereich, zb. genormte betriebliche oder behördliche Formulare

Mit dem Strukturierungsgrad steigt die semantische Spezialisierung sowie das Automatisierungspotential für diesen Dokumenttyp.

4.4 Anwendungsfelder

Tabelle 4.1: Automatisierte Dokumentenverarbeitung

Domäne	Primäre Dokumenttypen	Strukturierungsgrad	wirtschaftliche Relevanz
Buchhaltung	Rechnungen	2	++
Verwaltung	Formulare	3	+
Personal	Bewerbungen	1-2	o
Kanzleien	Schriftverkehr	1	+
Logistik	Lieferscheine	2	+
Privat	Rechnungen	2	-
	Versicherungen	1	Services für Privatanwender
	Formulare	3	nicht etabliert

4.5 Nutzungsmerkmale

Eine erste kurze Betrachtung der antizipierten Nutzungsmerkmale welche Entscheidungen zu Anwendungsdomäne (ref), Objektbereich (ref) und ... unterstützen soll.

Tabelle 4.2: Nutzungsmerkmale

Domäne Organisationsrolle	Arbeitsumgebung	Arbeitsgerät(?)
Buchhaltung Bürokm Fachkraft	Büro	Desktop PC
Verwaltung Bürokm oder Verwaltungs- Fachkraft	Büro	Desktop PC
Personal Personaldienstleistungskfm Fachkraft	Büro	Desktop PC
Kanzleien Bürokm Fachkraft	Büro	Desktop PC
Logistik Bürokm Fachkraft, Lagerist	Büro ggf Mobil, in größeren Betrieben	Desktop PC ggf Tablet
Privat	Zuhause Mobil	Desktop PC Tablet, Smartphone
Privat, Organisieren von Post(?)		

Kapitel 5

Alleinstellungsmerkmale

Die Markt- und Domänenanalyse offenbart (hoffentlich!!!) eine Marktlücke für Firmen mit

1. mit einem bestehenden Extrahierungsprozess und optionalem Nacherfassungsprozess
2. mit einem bestehenden Buchungs-, Kontierung- oder Archivsystem
3. Workflow
 - (a) mit bestehendem Workflowsystem oder
 - (b) ohne Bedarf für ein vollintegratives Workflowsystem
4. mit Bedarf für eine Automatisierung

Slim Clients

Automatisierte Attributierung

freie Wahl des Workflowsystem Für ein Workflowsystem müssen

Kapitel 6

Methodischer Rahmen

Kapitel 7

Kommunikationsmodel

7.1 Steuerungsclient

- erhält Information vom Verwaltungsdienst zum Systemzustand:
welche Absender aktuell sind wie lange im System
- teilt dem Verwaltungsclient (über Verwaltungsdienst) die Priorisierung mit:
nach Eingangsdatum (FIFO), Absender, ggf Anzahl/Absender

7.2 Verwaltungsclient

- erhält aktuelle Geschäftsobjekte vom Verwaltungsdienst
- gibt vervollständigtes Geschäftsobjekt an Verwaltungsdienst zurück
- gibt unvollständiges Geschäftsobjekt an Fachclient (über Verwaltungsdienst)

7.3 Fachclient

- erhält unvollständiges Geschäftsobjekt von Verwaltungsclient (über Verwaltungsdienst)
- gibt vervollständigtes Geschäftsobjekt an Verwaltungsdienst zurück

7.4 •

Kapitel 8

Risiken

8.1 Architektur

1. Paradigma das Lose Kopplung unterstützt
2. Steuerung der Priorisierung der Geschäftsobjekte -> im besten Fall unabhängig von Requests durch Clients

8.2 Technisch

1. Mobiler Client
 - (a) Kommunikation zu Steuerungskomponente im Verwaltungsdienst
 - (b) Implementierung von Interaktionsparadigmen (?)
- 2.

Kapitel 9

Proof of Concepts

9.1 Architektur

9.2 technisch: mobiler Client

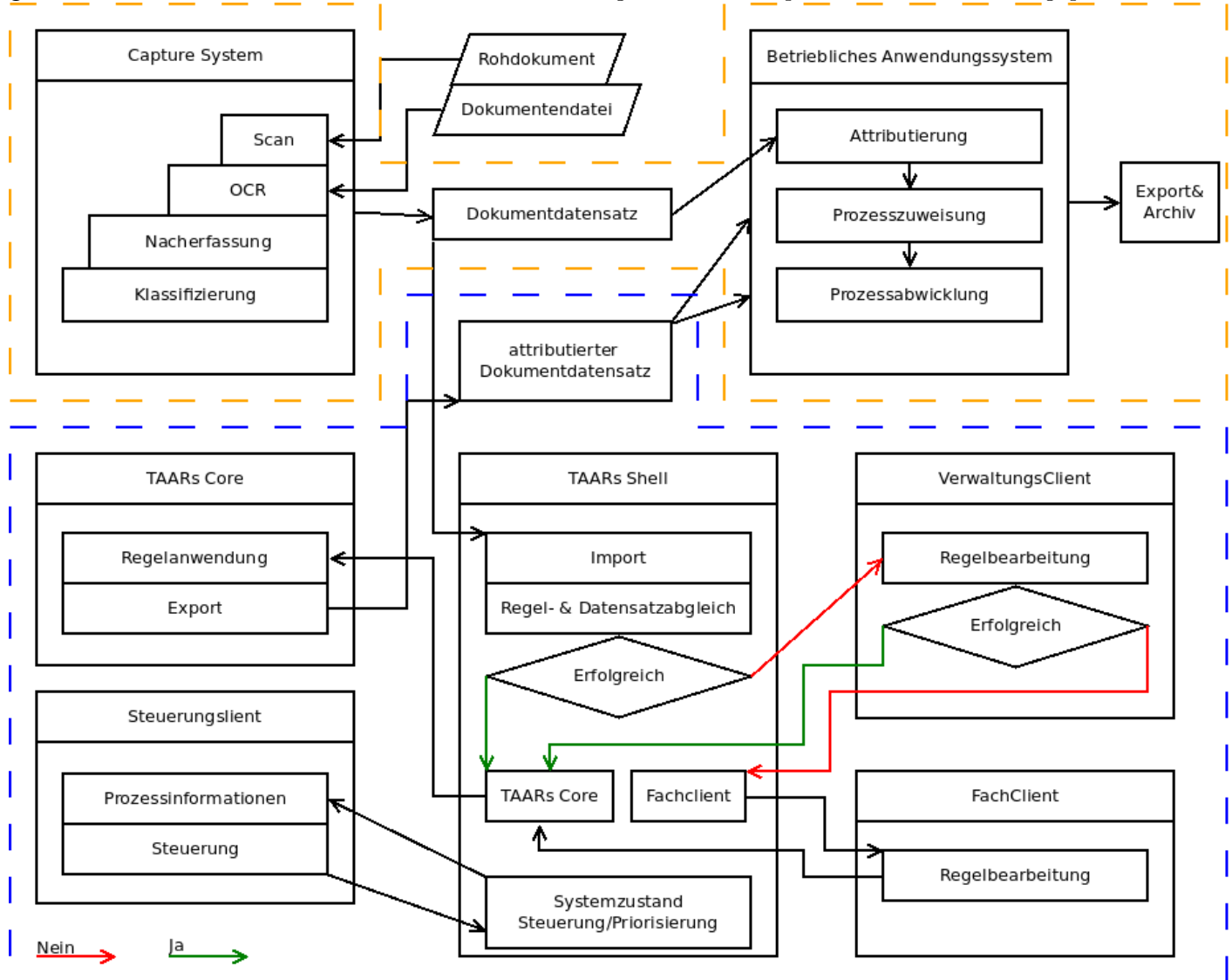
- ruft api des verwaltungsdienstes auf
- erhält json(?) objekt mit Informationen zum Systemzustand
- ruft Steuerungs api mit Priorisierungsparameter auf

9.3 technisch: Steuerungsaip Dienst

Kapitel 10

Prozess

Ausgehend von *Domaenenrecherche* ProzesseineDarstellungdeseinesexemplarischenVerarbeitungsprozessesunddieI



Kapitel 11

Architekturdiagramm

11.1 Übersicht

- *bsp arch*
- *arch mit ASys*
-

11.2 Verwaltungsdienst

11.3 Clients

11.3.1 Verwaltung

11.3.2 Fachclient

11.4 Steuerungsclient

11.5 Regel-Engine

Kapitel 12

Projektplan

Der Projektplan wird zunächst hier geführt.

Kapitel 13

Projektbegründungen

13.1 Implementierung

Die Entscheidung der Implementierungsumgebung wird aus den strategischen Zielen 1.1 sowie Punkt 3 der Kursziele abgeleitet, der da lautet:

Für die Bewerbungen in Unternehmen oder an Hochschulen ist heute oft neben einer guten Abschlussnote auch das Vorstellen einer anspruchsvollen, gut ausgeführten Projektarbeit ein wesentliches Erfolgskriterium. Das Praktikum hat das Ziel, den Studierenden die Möglichkeit zu geben, eine solche Arbeit zu erstellen oder zumindest einen ersten signifikanten Zwischenschritt bei Erstellung einer solchen Projektarbeit zu erreichen.

Daraus folgt die Erkenntnis das eine fachliche und technologische Annäherung des Projekts an den antizipierten beruflichen Kontext das Ausmaß der Zielerfüllung des Kurses erhöht

beruflicher technologischer Kontext *Im beruflichen Kontext wird für Windows Desktop und Windows Server im Stack .NET, c, MSSql entwickelt.*

Contra *Eine Entwicklung im og Kontext würde folgende Nachteile mit sich bringen:*

- 1. fehlende Unterstützung bei Implementierung durch Kursbetreuer*
- 2. fehlende Portierbarkeit der Komponenten*
- 3. ...*

Pro

- 1. höhere Bewegungssicherheit im beruflich relevanten technologischen Kontext*
- 2. Wettbewerbsvorteil durch Erwerb technologischer Kompetenzen 'abseits der Masse'*

Entscheidung *Daraus folgt die Entscheidung im beschriebenen technologischen Kontext zu implementieren. Es bleibt jedoch der Vorbehalt bei Bedarf einzelne Systemkomponenten in einem anderen Kontext zu implementieren.*

13.2 Objektbereich

Aus der ?? und der Punkt 1 und 2 der Strategisch folgt die Entscheidung das im Rahmnen dieses Projekts der Objektbereich auf den Dokumenttyp Rechnung eingegrenzt wird.

13.3 Nutzermodelle