

# EIS WS1516 Exposé

## Verteiltes Training einer automatisierten Dokumentenklassifizierung

Tim Howe

### 1 Nutzungsproblem

In Unternehmen werden digitale und analoge Eingangsdokumente, wie Rechnungen oder Lieferscheine, in einem Erfassungsprozess in eine vorbereitenden Struktur gebracht. Danach liegen die Dokumente als Dokumentendatei, Bilddatei sowie beschreibende xml-Datei vor und bilden die Grundlage für das zu verarbeitende Geschäftsobjekt. Abhängig von Dokumententyp und Ausprägung werden dem Geschäftsobjekt in weiteren Prozessen, zb. Buchung einer Rechnung, durch eine bürokaufmännische Fachkraft weitere Attribute zugewiesen.

Ein Großteil dieser Zuweisungen ist sehr repetitiv und variiert nur geringfügig. Daher wird dieser Prozess über eine Regel-Engine automatisiert, bei der die bürokaufmännische Fachkraft der Regel-Engine mitteilt welche Werte für eine Dokumentenausprägung zugewiesen werden sollen.

Es kann zudem vorkommen das diese Person nicht ausreichend fachbezogenes Wissen besitzt um eine Zuweisung durchzuführen. Dafür wendet diese sich an einen zuständigen fachkundigen Kollegen. Dies geschieht in der Regel über separate Kanäle (Telefon, EMail, ...), die ausserhalb des eigentlichen Prozesses verlaufen, was eine Reihe von Nachteilen mit sich bringt. Um diesen Nachteilen entgegenzuwirken wird ein Fachclient zur Verfügung gestellt an den ein Zuweisungsvorgang abgegeben werden kann.

### 2 Zielsetzung

- Verwaltungsclient: muss zügig entscheiden können ob er ein Geschäftsobjekt vervollständigen kann oder es an den Fachclient weitergibt
- Fachclient: sollte sich in Aufgaben und Abläufe des Fachangestellten wenig invasiv integrieren
- Regel-Engine: muss so trainiert werden können das der überwiegende Teil der Testobjekte mit einer zu definierenden Menge an Attributen automatisiert verarbeitet werden kann

### 3 Verteilte Anwendungslogik

- Präprozessor: Aufnahme der zugrunde liegenden Dokumentenstruktur, erzeugen des initialen Geschäftsobjekt, Prüfung, Weitergabe an Verwaltungsclient oder Regel-Engine
- Verwaltungsclient: Eingabe in das Regelsystem, ggf delegieren des Geschäftsobjekts an einen Fachclient
- Fachclient: unterstützung des Fachclients, fachliches vervollständigen der Eingabe
- Regel-Engine: Verarbeitung des Geschäftsobjekts anhand der trainierten Regeln, Rendern des zu exportierenden Geschäftsobjekts

### 4 Wirtschaftliche Relevanz

- (Teil-) Automatisierte Verarbeitung: Kostenreduktion, Vermeidung repetitiver Tätigkeiten, Befreiung intellektueller Kapazitäten der Akteure
- Vermeidung von Parallelkommunikation zur Aufgabenerfüllung: Reduktion kognitiver Belastung der Akteure und Fehleranfälligkeit bei Übertragung von Daten, zeitliche Ersparnis
- Slim-Clients: vereinfachte Integration in bestehende Prozesse ohne Zwang eines komplexen Document-Management- oder Workflow-Systems