

1 Domänenrecherche

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

1.1 Allgemeiner Verarbeitungprozess

Beispielhaft die wichtigsten funktionalen Komponenten des Dokumentenverarbeitungsprozesses beschrieben 2.2[1], wichtig bleibt anzumerken das in konkreten Implementierungen die Funktionalitäten verschwimmen können und keine klare Trennung wie hier beschrieben realisiert ist.

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.

1.2 Begrifflichkeiten

1.2.1 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.

1.2.2 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.

1.2.3 Geschäftsobjekt

Im Kontext der Anwendungsdomäne wird das zentrale verarbeitete Geschäftsobjekt aus einer Menge von Dokumentendatei und zugeordneten beschreibenden Daten gebildet. Das Dokument liegt dabei in der Regel als Repräsentation im Dokumentformat .pdf, als .tiff oder .jpg Bitmapdatei und die beschreibenden Daten als .xml Datei vor.

1.3 Anwendungsfelder

Tabelle 1: Automatisierte Dokumentenverarbeitung

| Domäne | Primäre Dokumenttypen | Strukturierungs- grad | 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 |

2 Nutzungsmerkmale

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

Tabelle 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 Privat, Organisieren von Post(?) | Zuhause Mobil | Desktop PC Tablet, Smartphone |

2.1 Speziellere Anwendungsdomänen

Wie unter anderen von Codia [1] beschrieben ergeben sich die Bedürfnisse in spezielleren Anwendungsdomänen wie zb Hochschulen nur teilweise durch spezielle Dokumenttypen wie eine Studentenakte. Auch hier spielen allgemeinere Dokumenttypen wie Rechnungen eine große Rolle.

2.2 Quellen

1. Wolfgang Riggert, Klassifikation & Extraktion, FH Flensburg abgerufen am 15.10.2015, url: Klassifikation & Extraktion
2. Codia DMS in Hochschulen, abgerufen am 15.10.2015, url: DMS in Hochschulen