



DocumentScan

Teknisk beskrivelse

DocumentScan – Teknisk beskrivelse

◆ Systemoverblik

DocumentScan er en Python-baseret applikation til automatiseret behandling af dokumenter (PDF/TIFF). Systemet arbejder batch-orienteret og er designet til at indgå i eksisterende dokumentflows.

◆ Arkitektur

- Filbaseret input/output
- Konfiguration via XML
- Valgfri databaseintegration
- Logging og fejlhåndtering

◆ Behandlingsflow

1. Input fra overvåget mappe
2. TIFF konverteres til PDF
3. Dokumenter scannes side-for-side for strekkoder
4. Dokument opdeles ved hver identificeret strekkode
5. Sider roteres ved behov
6. Output genereres med struktureret navngivning
7. Resultat logges (fil/database)

◆ Funktionalitet

Filhåndtering

- Batch processing
- Arkiv/fallback struktur

Strekkodebehandling

- Identifikation på tværs af sider
- Første forekomst anvendes som reference

Output

- Navngivning: <prefix><barcode>.pdf
- Separate filer pr. segment

Kvalitetsmåling (valgfri)

- Blur detection
- Kontrastanalyse

◆ Konfiguration

Via XML kan følgende styres:

- Input/output mapper
- Barcode placering
- Prefix-regler
- Databaseforbindelse
- Logging og notifikationer

◆ Drift

- Kan afvikles via scheduler (fx Windows Task Scheduler)
- Understøtter kontinuerlig batchafvikling
- Lavt memory footprint (sidebaseret behandling)
- Typisk rendering ved 300 DPI

◆ Integration

- Databaseopslag (prefix → metadata)
- Mulighed for kobling til ERP/DMS
- Klargøring til downstream-processer

◆ Udvidelser

Systemet er designet modulært og kan udvides efter behov.

Ændringer og integrationer specificeres og implementeres separat.