Extracting information from deeds by OCR and text interpretation

TS 2H – e-Land Administration

Rik Wouters Cadastre, Land Registry and Mapping Agency

Topics

- Introduction
- Extraction restrictions from deed
  - Business case
  - Basic steps
  - Techniques
  - Conclusions
Co-writers

- Rob Vaandrager Kadaster
- Gert Meijerink Unisys
- Jakub Zavrel Textkernel

The Netherlands Kadaster

Land Survey
Land Registration
Land Consolidation
Mapping
Information Services
Know your limitation

Extraction servitudes from deeds

Can I pass through here?
Do I allow people on my path?
Erfdienstbaarheden: Servitudes

- A servitude is a restriction on a real estate object through which the serving yard on behalf of the ruling yard is burdened. For example “Right of way”.
- Aim of the project scanning deeds and updating and making available the servitudes in the cadastral registration.
- Some 15,000,000 deeds from 1950 screened for
  - existence of servitudes and
  - bringing data in the cadastral registration

Business case

- To digitize 15 million deeds in the archives
- To extract restrictions from the deeds
- In a period of 4 years
- 20 M€ budget
- Use OCR and DTR techniques
- Benefits
  - Better service to notary
  - Centralized on line access to data
Archive of deeds

Project approach

• Develop process for
  – scanning
  – extraction
  – retrieval

• Research potentials of OCR and text recognition
  – Last attempts failed
  – Bad quality of originals
  – Start with 1950

• POC to test solutions
What do we do

Search each deeds for servitudes

Control of the found servitudes

Determine on which parcel the servitude confines

Digitize all deeds from 1950

Search each deed for servitudes

Determine to which parcel the servitude confines

kadaster
Text recognition

- Erfdienstbaarheid (-heden)
- Result of OCR
  - Erfdienstbaarheden,
  - lidjende erfdienstbaarheden
  - erfdienstbaarheden
  - erfdienstbaarheden,
  - erfdienstbaarheden,
  - erfdienstbaarheden,
  - erfdienstbaarheden
  - erfdienstbaarheden,
  - erfdienstbaarheden

Outcome of Text Recognition Search

- 3 possible outcomes: No, Yes or maybe
- > 55% of all deeds do not contain “erfdienstbaarheden” (99.8% sure)
- This implies that some 9,000,000 deeds do not need manual reading.
- In deeds that require manual reading:
  - all text parts that contain “Erfdienstbaarheden” are highlighted
- Tuning AI-software to improve results
What do we do

Digitize all deeds from 1950

Search each deed for servitudes

Checking of the found servitudes

Determine to which parcel the servitude confines

Control of found servitudes

TS 2H FIG Australia OCR and Text Recognition

FIG Congress 2010
Facing the Challenges – Building the Capacity
Sydney, Australia, 11-16 April 2010
What do we do

1. Search each deed for servitudes
2. Checking of found servitudes
3. Determine to which parcel the servitude confines

Questions can now be answered

Alle akten vanaf 1950 digitaliseren (Translate: All files from 1950 onwards digitized)

TS 2H FIG Australia OCR and Text Recognition
Project conclusions

- Project on schedule time/money
- Digital solution saves a lot of time
  - 55% of deeds does not need to be read with certainty of 99.8%
- Textkernel screens the deed better than a human being can do
- Information on servitudes on line is great help to Kadaster-staff and notaries
- Critical Success factors
  - Innovative Business Model: use of POC’s
  - Involve knowledge of private sector
  - Short lines between project teams
  - Create Win-Win situation (Partners in Business)
Thank you for your attention

More info: rik.wouters@kadaster.nl
www.oicrf.org