P14 - Extracting Names From Old Church Books

Maris Häusler
Egle Saks
Peter Kallaste
Daniel Nael
Project Owner

Contact person: Sulev Reisberg

Rahvusarhiiv
Church books

- Old books
- Contain lists of people’s names
  - Births
  - Marriages
  - Deaths
The Problem

- The need for getting the names and indexing them

- Doing it manually is time consuming and requires a lot of work
Our mission is to create a tool for identifying surnames from the microfilmed pages of church books automatically and create a surname index as a result.
Approach

- Created dataset for training
Approach

- Created dataset for training
- Trained a neural network

- CNN 5 layers
  - width = 128
  - height = 32
  - features = 256
  - time-steps = 32

- RNN 2 layers
  - time-steps = 32
  - characters = 80

- CTC loss
  - length <= 32
  - “Peipus”
  - 0.123...

- CTC decode
  - length <= 32
  - “Peipus”
Approach

- Created dataset for training
- Trained a neural network
- **Detected names using the model**
Approach

- Created dataset for training
- Trained a neural network
- Detected names using the model
- **Filtered out possible surnames**
Results

- Got the names (with flaws)
Lessons Learned

- Experience in detecting handwritten text
- Improved teamwork and communication skills
- Time management can be hard if you are not an expert in the topic
- Labelling takes a lot of time and work
Link to our Repository

https://github.com/dannynt/MLp14
Thank you for listening!