

4 - Creating Training Data from Existing Annotations

Background

- Loghi requires a specific training format
- each text line as a separate image
- index file:
 - path to line image + corresponding transcription
- one index for training & one for validation

Things We Need:

- path to training data
 - images
 - "page" folder: containing transcriptions in PAGE XML
- path in which to store reformatted training data

Running create-train-data.sh

```
./create-train-data.sh INPUT_TRAINING_DATA OUTPUT_TRAINING_DATA
```

Inspecting the Created Files