Download the IAM On-Line Handwriting Database

Structure

The IAM-OnDB is hierarchically structured into forms (The name of the files correspond to the naming scheme of the LOB Corpus):

- data/original-xml-part.tar.gz - Contains all forms which were acquired with the recording system of the University of Bern, i.e. the writer-ids and the transcriptions are stored within the xml-files. See a description of the xml-format.
- data/writers.xml - Contains the information of all writers in the database.
- data/lineStrokes-all.tar.gz - Contains the xml-files of the divided lines in on-line format.
- data/lineImages-all.tar.gz - Contains the xml-files of the divided lines in off-line format.
- data/original-xml-all.tar.gz - Contains all original forms. This is a superset of original-xml-part.tar.gz. It additionally contains forms with the same on-line handwriting informations but without the transcription and the writer-id. For that, see data/forms.txt, which contains the mapping of all forms to the writers.
- ascii-all.tar.gz - Contains the ascii-transcriptions of all files in the database (Under the CSR:-part).

You can download a compressed archive for each top level directory.
For a detailed information about the stored data see data.

Terms of usage

The IAM-OnDB is publicly accessible and freely available for non-commercial research purposes. If you are using data from the IAM-OnDB, we request you to register, so we are aware of who is using our data. If you are publishing scientific work based on the IAM-OnDB, we request you to include a reference to our database.

Handwritten text recognition task IAM-OnDB-t1

To make all the experiments on the IAM-OnDB comparable to each other in a quite adequate way, we ask you to use the setting as follows. Then you can refer to the recognition task as IAM-OnDB-t1:

- The database is divided into 4 parts, a training set, a first validation set, a second validation set and a final test set. The training set may be used for training the recognition system, while the two validation sets may be used for optimizing some meta-parameters. The final test set must be left unseen until the final test is performed. Note that you are allowed to use also other data for training etc, but report all the changes when you publish your experimental results and let the test set unchanged (It contains 3859 sequences, i.e. XML-files - one for each text line).
- Please use the characters of letters to be recognized and the whole dictionary found in spell, where also the spelling of the words is included. Note that there is an *sp* at the end of each word, since this is the official HTK notation of a spell-file. A word is correctly recognized only if all characters are recognized correctly (also in respect to upper case/lower case) and the boundaries are also found correctly. Here you can also change the letters to be recognized, but do not forget to mention these changes, especially if you have used a different dictionary or even an open-vocabulary dictionary.
- For the transcription of the text of all data in the task into character-level, please refer to labels.mlf.

You can also download all files in one tar.gz archive

Please report all the changes to the data/task you have made in your experiments. If your recognizer uses a language model, please refer to the texts on which this model has been trained. For a better comparison of the results, also a test without a language model would be interesting for the research community. Please state also these results, if applicable. The results of our recognition system will be published later on this website.

Handwritten text recognition task with open vocabulary IAM-OnDB-t2

This task is similar to IAM-OnDB-t1. There are changes in the vocabulary and letters file. This task can be refered to as IAM-OnDB-t2. Please use the following files:

- t2_letters (Please note, that there are more letters in this task.)
- t2_spell (Please note, that we added the small character a in front of each special symbol, because this is needed by HTK.)
- t2_labels.mlf