> **Attention**
>
> The site is currently being revised. If you encounter any problems, please **contact us.**

1.OCR4all

1.1 Introduction

OCR4all is a software that was developed for the digital text indexing, primarily of very early printed works, whose complex print types and often inconsistent layout concepts exceed the recognition capabilities of many other text recognition programs. The workflow proposed in OCR4all, which is understandable and can be used independently, also appeals to a group of users with a dedicated non-IT background and combines different work tools within a uniform user interface. In this way, constant switching between different programs is no longer necessary. From pre-processing of the image files to be processed (so-called pre-processing) to layout segmentation (so-called region segmentation with LAREX),

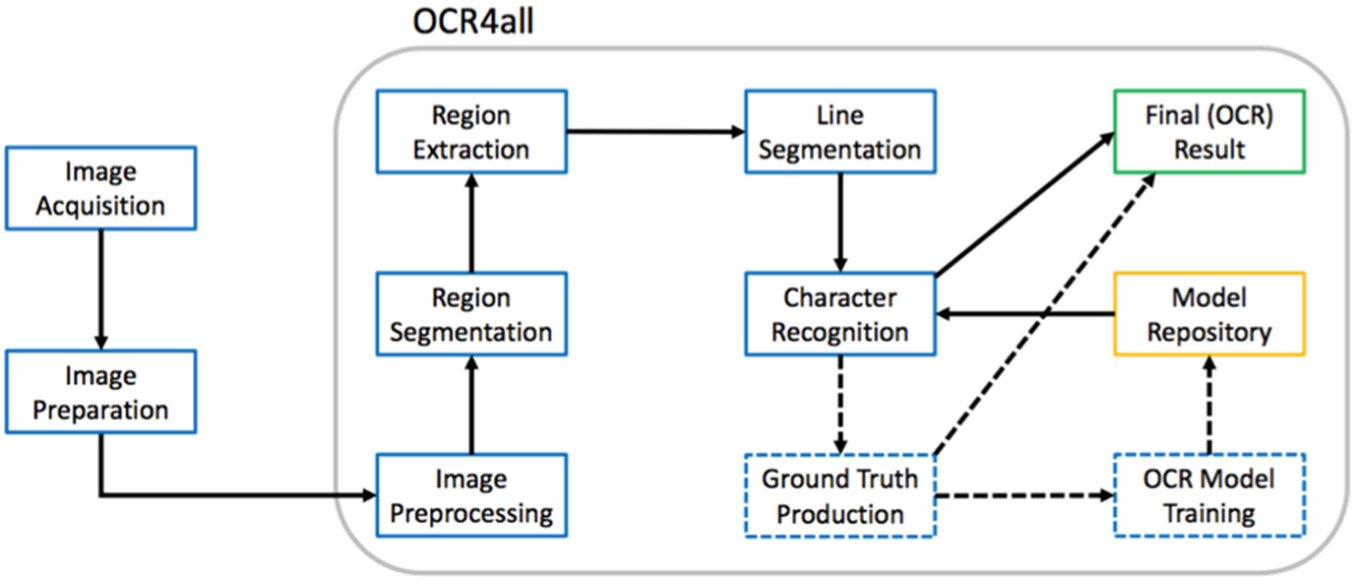

Fig. 1: Main components of the OCR4all workflow.

achieved in digital text indexing with OCR4all for almost all printed texts. With a view to this overall concept, these instructions provide a comprehensive and detailed insight into the work and possible uses of OCR4all in the context of the OCR of particularly early prints. While chapter 1 explains the general setup and folder structure, chapter 2 deals with a recommended pre-processing step of scans and image files outside of OCR4all, which improves results and simplifies the work steps inside OCR4all. Chapter 3 deals with starting OCR4all and an overview of the basic functions.The subsequent Chapter 4 then leads in detail through the different and successive sub-modules of the OCR workflow described above and presents the actual processing of prints and the creation of OCR texts in practice. The final chapter 5 is devoted to the currently most common user problems.

←0. Start

1.2 Setup and Folder Structure→