

GraphManuscribble 2.0

Vögtlin Lars

Master thesis in Computer Science

For digital libraries categorization and annotation of historical documents is important, as the produced Ground Truth is required for machine learning algorithms and quantitative evaluation of (semi-) automatic methods. This is important to perform layout analysis, text recognition or word spotting. Missing a reliable automatic method for creating Ground Truth a semi-automatic method, called GraphManuscribble, was introduced. That practice uses a document graph to interact with the document. Based on this graph, it gives the user the opportunity to produce fast, high quality Ground Truth in a modern fashion. But GraphManuscribble works on a pixel based representation of document graphs and thus produces long waiting times and provides no direct feedback for the user. In this thesis we present a new version of the GraphManuscribble reducing these drawbacks. By analyzing and adapting the used algorithms, we were not only able to create a vector based version of GraphManuscribble, but also to reduce calculation time, mainly of the concave hull algorithm, and improve user-friendliness.

Prof. Rolf Ingold