

PRIP-TR-18

February 12, 1993

Layoutanalysis

Finding Text, Titles, and Photographs
in Digital Images of Newspaper Pages

Diploma Thesis Goals Report

Johann Wieser

Abstract

An important step in the analysis of printed documents is the segmentation and classification of blocks into categories such as photographs, titles, paragraphs, etc. This report presents an approach to enhance and combine two commonly used methods, a merging bottom up approach and a cutting top down approach, to segment pages of a newspaper. The planned procedure for an implementation of a layoutanalysis system as a preprocessing module for a commercial product is described.