

Předmětem této práce je navrhnout zpracování skenovaných dokumentů obsahujících textové a netextové elementy. V této práci je tento problém analyzován a rozdělen na menší podproblémy, které jsou dále řešeny. Hlavní řešené problémy jsou předzpracování obrázku, při kterém se mají potlačit nedostatky skenování, například špatný kontrast či našikmo sejmутý obraz, dále je řešena segmentace dokumentu na znaky, linky a obrázky a závěrem komprese výstupu pomocí clusterování znaků. Dále je cílem této práce navržené řešení implementovat formou programu, který naskenovaný dokument zpracuje a vytvoří PDF dokument skládající se z nalezených komponent.