V této práci jsem vyvinul aplikaci schopnou extrahovat data z naskenovaných dokumentů. Pro optické rozpoznávání znaků jsem použil externí OCR engine Tesseract, který lze snadno vyměnit. Pro jednotlivé doklady používám šablony s informacemi o datových oblastech a jejich datových typech. Pokusil jsem se automatizovat většinu kroků nutných pro extrakci dat a vytvoření nové datové šablony. Uživatel má možnost opravit nebo změnit výsledky těchto kroků. Pro výstup z aplikace jsem implementoval komponenty, které exportují data do formátů XML, HTML a do obyčejného textu. Další komponenty mohou být snadno přidány, aby přizpůsobily aplikaci různým použitím.