

Avaliação da eficácia da extração de texto em imagens

Daniel Matheus Kuhn, Edimar Manica, Cristiano Roberto Cervi*

Reconhecimento Óptico de Caracteres (OCR - Optical Character Recognition) consiste em tecnologias capazes de interpretar caracteres em imagens textuais ou mapas de bits. Em sua origem foram aplicados, principalmente, em arquivos escaneados. O advento de dispositivos com câmeras de alta qualidade abriram novas oportunidades para aplicações. A Google Inc, por exemplo, vem empregando OCR em imagens coletadas por meio de seu serviço de mapa Google Maps, com o objetivo de extrair e indexar elementos textuais de placas de trânsito, números residenciais, bem como fachadas de lojas, transformando em informações pesquisáveis em seu motor de busca. O presente trabalho tem como objetivo avaliar a eficácia da extração de textos em imagens de trechos de livros capturadas por smartphones. Para a realização do experimento, foi selecionado um grupo de 6 (seis) usuários. Os usuários capturaram 2 (duas) imagens de trechos de livros. Cada imagem foi submetida isoladamente para o framework de reconhecimento de caracteres Tesseract. Esse framework é totalmente treinável, possibilitando que sejam incluídas a ele, novas fontes, símbolos ou até mesmo idiomas inteiros. Além disso, Tesseract é de código aberto, ou seja, seu código fonte pode ser estudado, modificado e distribuído sem custo para qualquer usuário. Para realizar a avaliação da eficácia da extração realizada pelo Tesseract, o texto extraído pelo framework foi comparado com o gabarito. O gabarito contém o texto de cada imagem, que foi manualmente identificado por um usuário especialista. Foram adotadas as métricas de precisão, revocação e F1, consideradas tradicionais na área de recuperação de informação. A precisão mede a fração das palavras recuperadas que é relevante. A revocação mensura a fração das palavras relevantes que foi recuperada. Uma palavra relevante é aquela que foi extraída de uma imagem corretamente e, portanto, está presente no gabarito. A F1 combina a precisão e a revocação usando uma média harmônica entre as métricas. Em média, o Tesseract atingiu uma revocação de 50,35%, uma precisão de 51,39% e uma F1 de 50,23%. Tesseract apresenta resultados satisfatórios quando aplicado a imagens com perspectiva frontal do texto em relação ao smartphone e, com pouco grau de inclinação e curvatura das linhas do texto. Entretanto, há uma perda expressiva nos indicadores de eficácia em situações adversas de perspectiva e inclinação. Isto se deve ao fato de o Tesseract não implementar pré-processamentos de rotação e correção de perspectiva. Para trabalhos futuros, orienta-se dar ênfase a esses aspectos, adicionando etapas de pré-processamento das imagens.

Palavras-chave: *OCR. Extração. Imagens textuais. Tesseract.*