OCR Demo

# Visao Geral

OCR é um acrónimo para o inglês Optical Character Recognition, é uma tecnologia para reconhecer caracteres a partir de um arquivo de imagem ou mapa de bits sejam eles escaneados, escritos a mão, datilografados ou impressos. Dessa forma, através do OCR é possível obter um arquivo de texto editável por um computador.

# Tecnica Utilizada

Podemos citar vários métodos para extração e reconhecimento desses caracteres. Para esta demonstração foquei em utilizar a engine de código aberto Tesseract que é considerado como provável mais preciso open source de OCR.

Os dados treinados utilizei o inglês, pois como a ferramenta foi desenvolvida em inglês o resultado apresentou maior precisão, sendo assim utilizá-lo para textos em português pode haver problema no reconhecimento de caracteres exclusivos do nosso idioma (e.g.: ç, á, ã , etc).

# APK

Dentro da pasta *OCRdemo/app/* tem o arquivo .APK, instalador do app *OCRdemo-release.apk.*

# Referencias

**Tesseract-OCR**

https://code.google.com/p/tesseract-ocr/

**Tesseract-OCR - Language Data**

https://code.google.com/p/tesseract-ocr/downloads/list

**TessBaseAPITest**

https://code.google.com/p/tesseract-android-tools/source/browse/tesseract-android-tools-test/src/com/googlecode/tesseract/android/test/TessBaseAPITest.java

**Wikipedia**

https://pt.wikipedia.org/wiki/Reconhecimento\_%C3%B3tico\_de\_caracteres

# Contato

Desenvolvedor: Tiago Oliveira

E-mail: tiago\_fernandes89@hotmail.com