

OCR para extração de texto em mangás

Bryan Jonathan Melo de Oliveira

Daniel Soares França

Eros Henrique Rodrigues Oliveira do Monte

Tallys Assis de Souza

Introdução

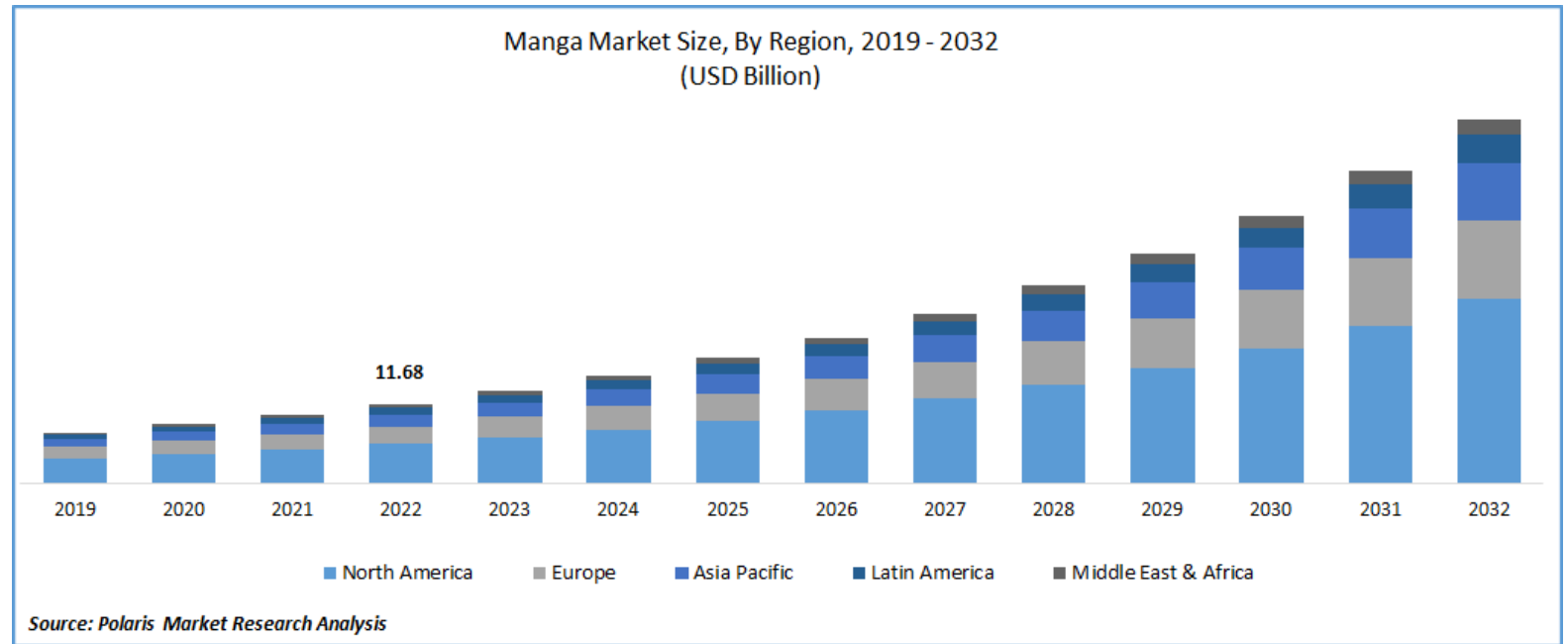
Uma grande quantidade de mangás são produzidos todos os dias

Muito tempo e esforço são necessários para a tradução

Produzir uma aplicação para a extração de texto de mangás

Reduzir o custo necessário para a tradução e aumentar a produtividade

Mercado



Estima-se um mercado de 50 bilhões em 2030

Poucas editoras trabalham na área

Revisão literária

An Overview of the
Tesseract OCR Engine - by
Ray Smith

Improving the Accuracy of
Tesseract 4.0 OCR Engine
Using Convolution-Based
Preprocessing - by Dan
Sporici, Elena Cuşnir and
Costin-Anton Boiangiu

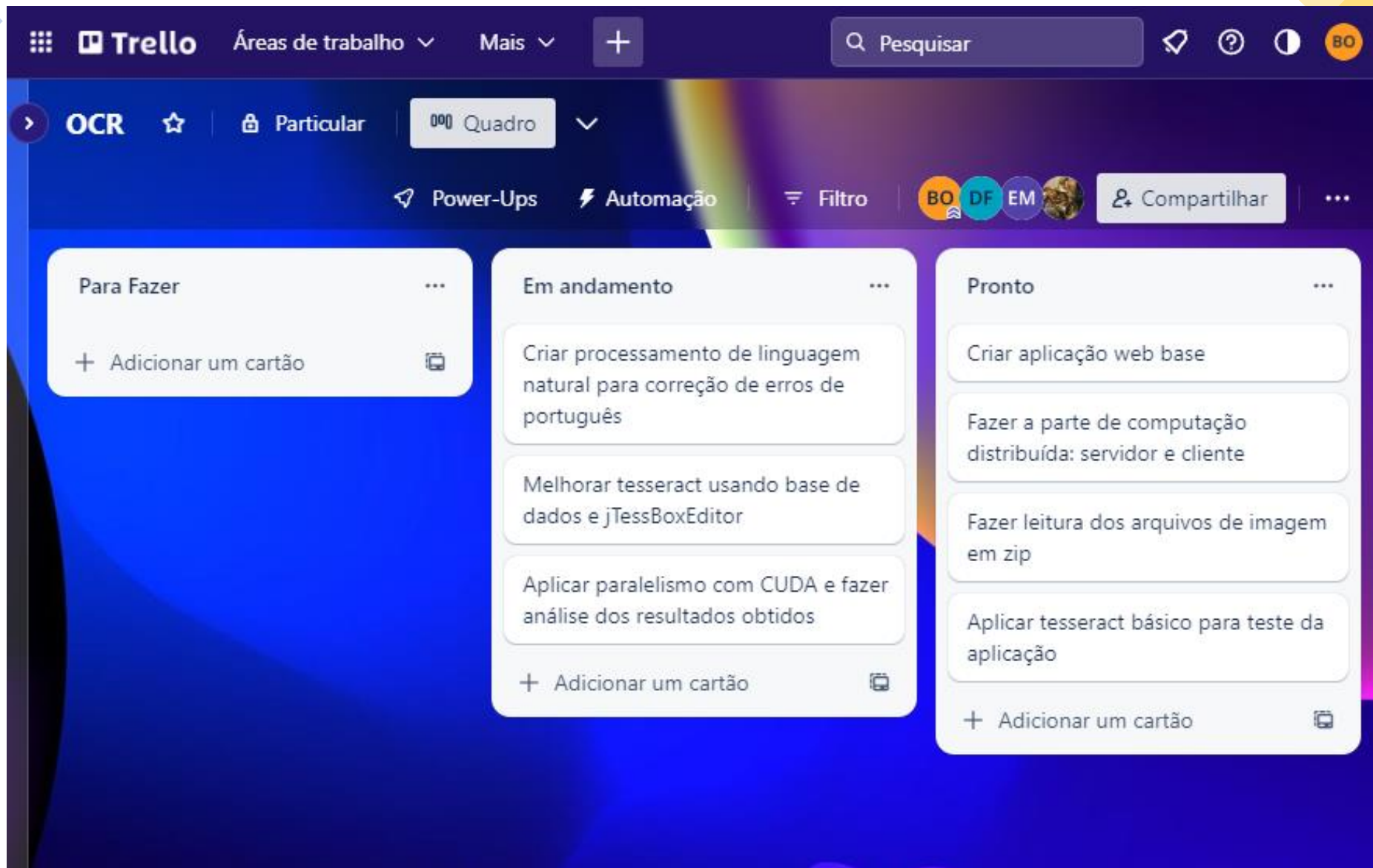
Avaliação e metodologia

Para a distribuição de tarefas: Trello

Para a parte paralela da aplicação: CUDA

Optical character recognition: Tesseract

Para a parte de computação distribuída: Cliente-Servidor

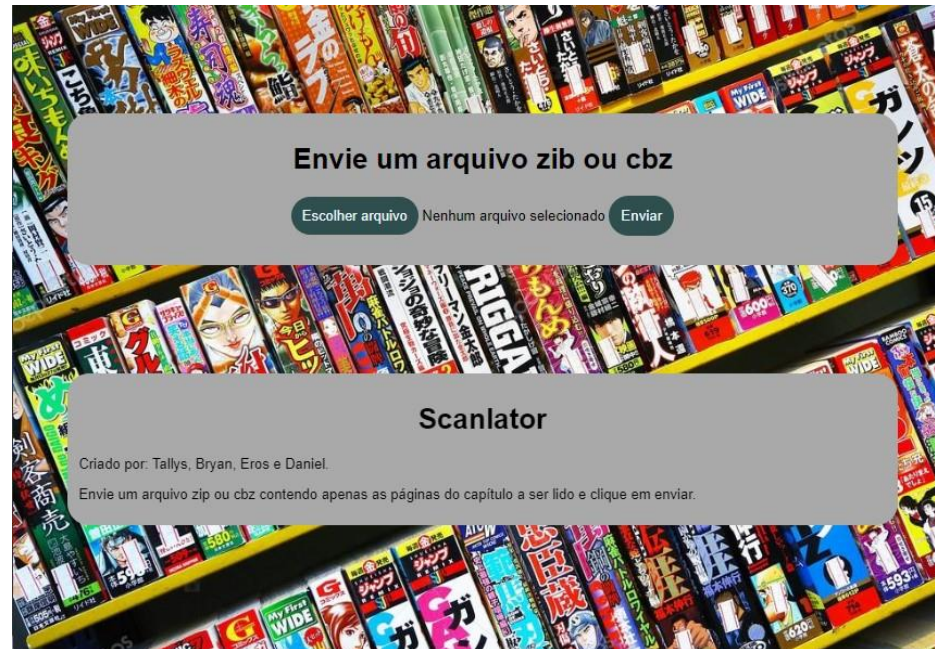


Considerações Iniciais

Falta de trabalhos relacionando
optical character recognition e
mangás

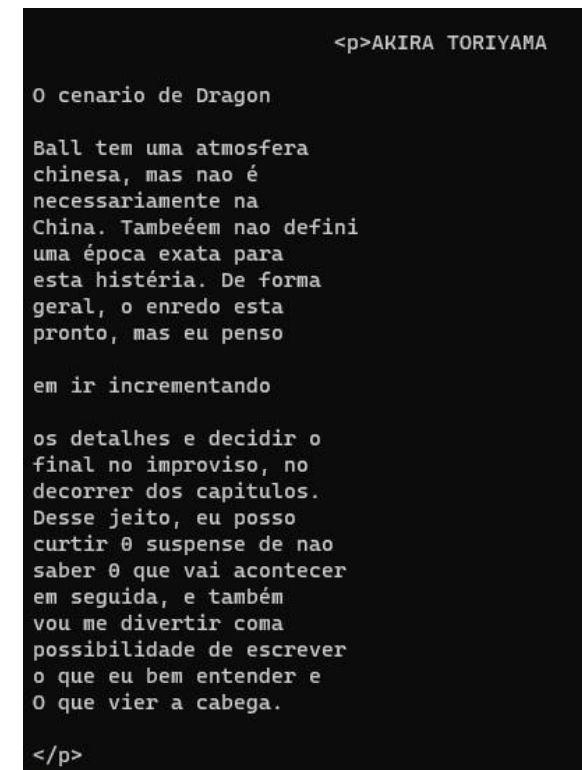
Acreditamos que será necessário a
aplicação de processamento de
linguagem natural para a correção
de erros de português

Resultado inicial



Results

AKIRA TORIYAMA O cenário de Dragon Ball tem uma atmosfera chinesa, mas não é necessariamente na China. Também não defini uma época exata para esta história. De forma geral, o enredo está pronto, mas eu penso em ir incrementando os detalhes e decidir o final no improviso, no decorrer dos capítulos. Desse jeito, eu posso curtir o suspense de não saber o que vai acontecer em seguida, e também vou me divertir com a possibilidade de escrever o que eu bem entender e o que vier à cabeça.



Referências

- <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/33418.pdf>
- <https://www.mdpi.com/2073-8994/12/5/715>