

OCR para extração
de texto em mangás

Introdução

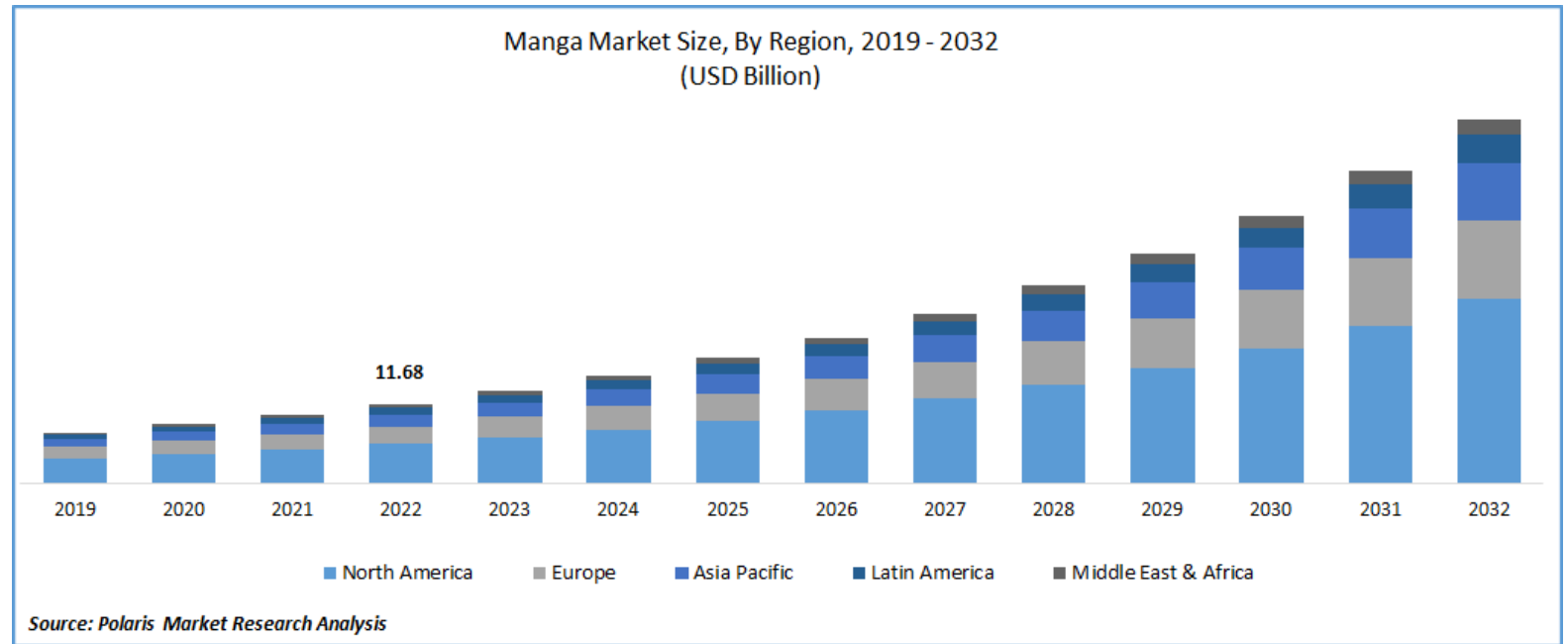
Uma grande quantidade de mangás são produzidos todos os dias

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

Possuimos um código base pra parte de cliente-servidor que tinha sido produzido para uma tarefa de computação distribuída

Parte do processamento feito pelo OCR já está funcionando

Referências

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