

# Aplicação de análise morfológica para segmentação de páginas em imagens de documentos

Aluno: Ricardo de Cillo  
Supervisora: Nina S. T. Hirata

## 1 Objetivo

Construir um software que receba como entrada a imagem de um documento digitalizado e produza como saída uma descrição geométrica [1] [2] de suas regiões de texto. Este software será treinado com exemplos de entrada e saída desejadas.

## 2 Metodologia

O problema de análise de layout de um documento possui duas principais abordagens:

**Bottom-up** parte do menor elemento na imagem (pixel), classificando-os e agrupando-os para formar as regiões. Exemplo: [3].

**Top-down** trabalha com elementos maiores que o pixel para inferir a estrutura do layout. Exemplo: [4].

Neste trabalho exploramos a aplicabilidade de operadores morfológicos como ferramenta para segmentação de documentos. Dado que a construção de operadores morfológicos pode consumir muito tempo, requerir muita experiência e gerar resultados suscetível a erros, utilizaremos um método para geração automático de operadores morfológicos a partir de imagens de treinamento [5].

## Referências

- [1] 10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009. IEEE Computer Society, 2009.
- [2] A. Antonacopoulos and D. Bridson. Performance analysis framework for layout analysis methods. *Document Analysis and Recognition, International Conference on*, 2:1258–1262, 2007.
- [3] Michael A. Moll and Henry S. Baird. Document content inventory and retrieval. In *In Proc., IAPR 9th Int'l Conf. on Document Analysis and Recognition (ICDAR2007*, 2007.
- [4] Raymond W. Smith. Hybrid page layout analysis via tab-stop detection. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, ICDAR '09*, pages 241–245, Washington, DC, USA, 2009. IEEE Computer Society.
- [5] N. S. Tomita. Programação Automática de Máquinas Morfológicas Binárias baseada em Aprendizado PAC. In *Anais do XXIII Seminário Integrado de Software e Hardware - IX Concurso de Teses e Dissertações - XV Concurso de Trabalhos de Iniciação Científica*, pages 517–525, Recife, PE, 1996. XVI Congresso da SBC, Sociedade Brasileira de Computação.