



Tesseract Blends Old and New OCR Technology

0. Tutorial Contents

Ray Smith, Google Inc.

Tutorial Contents

1. 14:30-15:00 Introduction/Motivation/History
2. 15:00-15:05 Architecture and Data Structures
3. 15:05-15:30 Features and Character Classifier
4. 15:30-15:45 Character Segmentation, Language Models & Beam Search
5. 15:45-16:00 Page Layout Analysis
6. 16:00-16:30 Break
7. 16:30-17:00 Modernization Efforts
7. 17:00-17:30 Building a Multilingual OCR Engine