Tessetact OCR 引擎总览

Ray Smith

Google Inc.

摘要

Tesseract OCR 引擎，过去是惠普的研究样品，在内华达大学（UNLV）第四次年度 ORC 准确性测试中被全面的描述。（测试的）重点被放在一个新颖或者至少不寻常的 OCR 引擎上，特别是包括了线性搜索，特征／分类方法和自适应分类器。

1.介绍 - 动机和历史

Tesseract 是一个开源的 OCR 引擎，它由惠普公司于1984年到1994年间所开发。在内华达大学（UNLV）1995年的 OCR 精度的年度测试中，它的测试结果像一个超新星那样迸发出光芒并且随后，它被