

أتمتة بناء التعرف الضوئي على الحروف (OCR) بناء على خدمات الويب

إعداد

لطفيه بنت سعود بن عبدالعزيز الحميد

إشراف

أ.د. كمال منصور سعيد جمبي

المستخلص

التعرف الضوئي على الحروف (OCR) هو نظام كمبيوتر مصمم لتحويل صور النص المكتوبة (عادةً ما يتم التقاطها بواسطة ماسح ضوئي) إلى نص معالج آلياً. وقد تم تطوير وتعزيز OCR العربية على مدى عقود مما أدى إلى وجود عدد كبير من الطرق مع نتائج قوية قريبة من الدقة في بعض الحالات ما يقارب ٩٩ ٪. ومع ذلك ، فإن معرفات OCRs الحالية تظهر نقصاً أو على الأقل مجموعات فرعية منها عندما يتم تنفيذها بتطبيقات جديدة مثل المدخلات منخفضة الدقة.

في هذه الأطروحة ، سنحسّن نتائج التعرف الضوئي على الحروف (OCR) عن طريق اختيار صورتنا الأصلية بناءً على قاعدة بيانات النص العربي المطبوع (APTI). بعد ذلك ، سنقوم بتحسين المدخلات من خلال معالجة الصور لجعلها واضحة وجاهزة للإدراج في OCR. ثم سنقوم باختبار جميع مواقع OCR باللغة العربية على الإنترنت للعثور على أفضل موقع موجود حالياً. وأخيراً ، سننشئ خدمة ويب يمكنها تحسين تأثيرات الصور تلقائياً وإدخالها في أفضل OCR العربية لتحقيق تحويل OCR العربي الأمثل. من جهة أخرى ، يوفر نظامنا قابلية الاستخدام من قبل أشخاص آخرين على الويب.

Automating Building of Optical Character Recognition (OCR) Based on Web Services

By
Lutfieh Saud Abdulaziz Alhomed

Supervised By
Prof. Dr. Kamal M. Jambi

ABSTRACT

Optical Character Recognition (OCR) is a computer system designed to transform images of typewritten text (typically captured by a scanner) into machine-processed text. Arabic OCR has been developed and enhanced over decades leading to the presence of the enormous number of approaches with robust results that approaching accuracy, in some cases, of approximately 99%. However, existing OCRs exhibit shortages, or at least only sub-sets of them, when they were implemented with new applications, such as low-resolution inputs.

In this thesis, we will improve results Optical Character Recognition (OCR) by choose our original image based on (APTI) Arabic Printed Text Image Database. Next, we will improve inputs by processing the images to make them clear and ready to be inserted into the OCR. Then, we will test all Arabic OCR websites online to find the best one which currently exists. Finally, we will create web service that can optimize image effects automatic and insert it to best the Arabic OCR to make the

perfect Arabic OCR conversion. In the second hand, our system available for use by
other people on the web.

