

ERA

“مشروع سوفتوير”

هو أحد تقنيات معالجة الصور الرقمية المستخدمة "Optical Character Recognition OCR" التعرف البصري على الحروف و الرموز للتعرف على الرموز و الحروف الموجودة على الصور و استخراج هذه الرموز لاستخدامها بعد ذلك لأي غرض و بالصيغة المناسبة. هذه "Machine" التقنية تعتمد على معالجة الصور قبل استخدامها و استخراج الشوائب منها و أيضا على الذكاء الصناعي و تدريب الآلة للحصول على نتائج أكثر دقة، و لتحقيق ذلك نحن بحاجة لعدد كبير من البيانات ليتم تدريب التطبيق عليها حتى يسهل عليه "Learning التعرف على الرموز من خلال اي صورة تعطي له

و هو تقنية قائمة على التعرف على الرموز و "Tesseract" واحد من أهم التقنيات المستخدمة في التعرف على الرموز هو التزراكت استخراجها من الصور النقية الخالية من الشوائب ذات الخلفية البيضاء، بحيث يتم استخدام هذه التقنية في عملية تدريب الجهاز على التعرف على نصوص محددة، بعدها يتم إنتاج ملف يحتوي على كل بيانات التدريب لنصوص معينة و يمكن استخدام هذا الملف في التطبيق أو الموقع الخاص بك للمساعدة في عملية استخراج النصوص من الصور بعد تنقيتها

تم بناؤه ليتعرف على رموز اللغات القديمة، و هذه الرموز وجدت منقوشة على الصخور و الخامات الأخرى و لا يوجد أي "ERA" مشروع مصدر أو تطبيق يدعمها أو يمكنه التعرف عليها، لذلك عملنا في هذا المشروع على دعم 3 لغات قديمة غير مدعومة من أي مصدر و هي العربية الجنوبية القديمة و العربية الشمالية القديمة و النبطية و بإمكانه دعم لغتين أخرتين و هما الرومانية و الإغريقية القديمتان المدعومة من مصادر أخرى، حيث تم توفير عدد كبير من البيانات كصور ليتم استخدامها كمدخلات في عملية التدريب على الثلاث لغات. و بما أن المشروع يدعم هذه اللغات الموجودة كمنقوش على الصخور يجب على المستخدم أخذ صورة للنصوص فيقوم التطبيق بتنقيتها و إزالة الشوائب منها ثم يقوم باستخراج الرموز منها على شكل نص ليتم استخدامه فيما بعد، و يوفر التطبيق تقنية الرسم على الصور إذا لم يتم التعرف على النصوص بالشكل الصحيح، فيمكن للمستخدم الرسم فوق الصورة لمساعدة التطبيق في التعرف على الرموز و الحروف