

تقييم تأثير استخدام أسلوب البرمجيات المتعددة على دقة برمجيات التعرف الضوئي على الحروف العربية

"دراسة تطبيقية"

يوسف بن عبد الحليم بناوي

المشرف الرئيسي على الرسالة

الأستاذ الدكتور أسامة بن احمد أبو النجا

المستخلص

في الماضي كانت معظم المؤسسات والشركات تعتمد على المعاملات الورقية، ومع مرور السنوات أدى ذلك إلى زيادة أعداد المعاملات الورقية في الأرشيف، فأصبح الوصول إلى المعاملة المطلوبة تستغرق وقتاً وجهداً. مما أدى إلى ظهور الأرشفة الالكترونية التي أصبحت امراً ضرورياً لمعظم الشركات. ومع التطور التقني في مجال الحاسب الآلي والرغبة في زيادة التفاعل بين الإنسان والحاسب الآلي صممت برمجيات التعرف الضوئي على الحروف. حيث يقوم البرنامج بتحويل صورة النص المأخوذة بالماسح الضوئي إلى مستند نصي يمكن للمستخدم التعديل عليه مما يوفر الوقت والمال والجهد للمستخدم.

وتعتبر دقة برمجيات التعرف الضوئي للحروف للغة العربية ليست مرضية نظراً لتعقيد حروف اللغة العربية. ولذلك يوجد عدد قليل من الشركات التي تُطور في هذا المجال.

ويعتبر أسلوب البرمجيات المتعددة أحد أكثر الأساليب المستخدمة للتعامل مع إخفاقات الحاسبات الآلية. حيث يستخدم هذا الأسلوب عدد من البرمجيات المختلفة ومن ثم اخذ الناتج حسب الأغلبية.

من خلال هذا البحث سوف ندرس تأثير استخدام أسلوب البرمجيات المتعددة على دقة مخرجات برمجيات التعرف الضوئي على الحروف العربية للنصوص المطبوعة. وتتضمن الدراسة برمجيات التعرف الضوئي على الحروف العربية التالية: القارئ الآلي لشركة صخر النسخة الذهبية الثامنة (Sakher Gold Ver.8) ، القارئ الآلي لشركة أيريس نسخة الشرق الأوسط الحادية عشر (Readirs Ver.11 Middle East) والقارئ الآلي فاجي شيناس النسخة الأولى (Vajeh Shenas 1st edition) .

Evaluating the Effect of N-Version Programming Technique on Arabic OCR Accuracy: Experimental Study

**By
Yusof A. Batawi**

Thesis Advisor

Prof. Osama A. Abulnaja

Abstract

In the past, most companies and organizations rely on paper-transactions. Over the years, this lead to the increased number of papers in the archive which make the access to the desired information take time and effort. This leads to the E-Archiving concept which becomes very essential to many companies. With technical evolution in the area of computer and the desire for greater interaction between human and machine; software such as Optical Character Recognition (OCR) were designed. OCR software transfers a scanned image to an editable text file. This process saves time, money and effort.

The accuracy of Arabic OCR (AOCR) software is not satisfying due to the difficulties to recognize the Arabic characters. Few companies develop AOCR software.

N-Version Programming (NVP) technique is one of the most widely used software fault-tolerant techniques. NVP depends on the redundancy of the programs and the voter to produce a majority output.

In this research we have studied the effect of using NVP scheme on the accuracy of the AOCR software. AOCR software that are included in the study are: Sakher Gold Ver.8, Readirs Ver.11 Middle East, and Vajeh Shenasa 1st edition.