

تقصينا

أهم الشركات التي تقدم برامج للتعرف على

المحارف العربية ووجدنا أن شركة Nuance تقدم عدة تطوير برمجية Omnipage Capture SDK، وهي مصممة للمطورين من الشركات التي تقدم حلول التعرف على المحارف (تجرب لنا لاحقاً أن صخر تستخدم هذا الحل). وكما وعداكم عقب معرض جيتكس بتقديم تقييم عملي للثلاثة برامج للتعرف على المحارف العربية فيها نحن نعرض النتيجة، لنضع النقاط على الحروف في لغة برامج التعرف على الحروف العربية، ومن يسعى إلى العثور على عبارة أو حكم مطلق لإقضاء برنامج ما فوز برنامج آخر لسبب ما لأنه لأن البرامج الثلاثة تتدفع بمزايا جيدة وتشارك كل برامج المجموعة التي قمنا بتقييمها واختبارها، في توافرها بعدة إصدارات منها للتصوير المكتبي الفردي وأخرى للتخام وأخرى تضم عدة تطوير SDK Edition. تنفيذ تكامل مع برامج وأنظمة الشركات القائمة لديها. لا نحظى قول النتيجة لكنها وجدنا أنها مناسبة، أي أن الحكم النهائي يتوقف على طبيعة الاستخدام المطلوبة. وهناك إشارة لعمل الدقة في كل منها، ورغم تفوق برنامج فيروس في الدقة إلا أن ذلك لا يعني أنه الفائز المطلق فعامل الدقة على أهميته لا يكفي دون مزايا أخرى. يتميز برنامج فيروس Verus من شركة نوفودايميكس للتعرف على المحارف العربية بفضل ميزتين أساسيتين أولهما الجودة في التعرف على أكبر عدد متوفر من الخطوط العربية المختلفة بأنواعها، وثانيهما القدرة على تخطيط وزيادة وضوح الوثائق من خلال محرك خاص لتلك الناحية بالأخذ بعين الاعتبار طبيعة اللغة العربية وخصوصيتها، مؤمن برامج أيريس عملية تحويل تلقائية للنص من الصور والوثائق المضمومة إلى البنية الرقمية في ملفات قابلة للبحث والتحرير والإضافة وغير ذلك من وظائف مفيدة، فمن أرشفة الشبكات رقمياً وتعبئة سجلات الفواتير إلى تعاملات أخرى عديدة مثل فرز الرسائل والطرود والحفاظ في الطارات، يمكن استخدام هذه البرامج للوصول إلى كفاءة عالية لدى الشركات والجهات الحكومية والبنوك وغيرها. يمثل نطاق تحويل النص المطبوع والتعرف على الحروف والرموز العربية تحدياً تقنياً كبيراً يستخدم فيه طرائق متعددة للوصول إلى دقة أكبر لا يوجد حضان أبيض وآخر أسود أو فائز يتحسر محقق في مجموعة اختباراتنا لبرامج التعرف على المحارف العربية. إذ بينما يتفوق منتج على البقية في ناحية محدد فإن الأخران يبرزان بمزايا أقوى في نواح أخرى. ومثلاً إذا أريد توليد ملفات إلكترونية بنسق PDF وقدرات ضغط الصور والتعرف على المحارف، فضلاً عن دعم أجهزة المسح الضوئي الملونة التي تسمح الوثائق على الوجهين (بيلكس) مع قدرات فصل وفهرسة الوثائق والتعرف على الجدول فيها، وإزالة الوثائق والأنظمة الاستمارات FORMS AUTOMATION وفي إدمه الشبكات لأن منتج أيريس يبرز بقوة. لقماً باستخدام عدة عميات من المستندات منها مستندات نظيفة ومنها ممسحة ونقص بالمستندات الممسحة تلك التي تحتوي على تشوهات أو على مزيج لوني معقد يجعل من الصعب التمييز بين الخلفية والنص. والكم النتائج بالأرقام

برامج التعرف على الأحرف العربية

Readiris

Readiris 11 Pro Middle East Corporate Edition
السعر ٩٩٩ دولار
الاتصال +٩٧١٤٢٨٢٨٦٢
الموقع www.inslink.com

ما ينفرد به ريدأيريس عن بقية المنافسين هو دعمه لأكثر من ١٠٠ لغة، واحتواءه على شريط أدوات يساري يمكن من إجراء العديد من التحكمات على المستند مثل التدوير أو التكبير والتصغير أو تحديد منطقة محددة من الصورة للتعرف على الأحرف التي تحتويها لزيادة الدقة.

يوفر هذا البرنامج إمكانية التعليم (أي تعليم البرنامج يدويًا على بعض الحروف) ليتمكن من التعرف عليها في المرات المقبلة بالشكل الصحيح. وهذه الميزة موجودة أيضًا في برنامج صخر، ويدعم البرنامج التعرف على عدد كبير من المساحات الضوئية بشكل مسبق. يسجل لبرنامج ريدأيريس احتواءه على معالجات لمساعدة الأشخاص المبتدئين في عالم التقنية على التألف معه بسرعة.

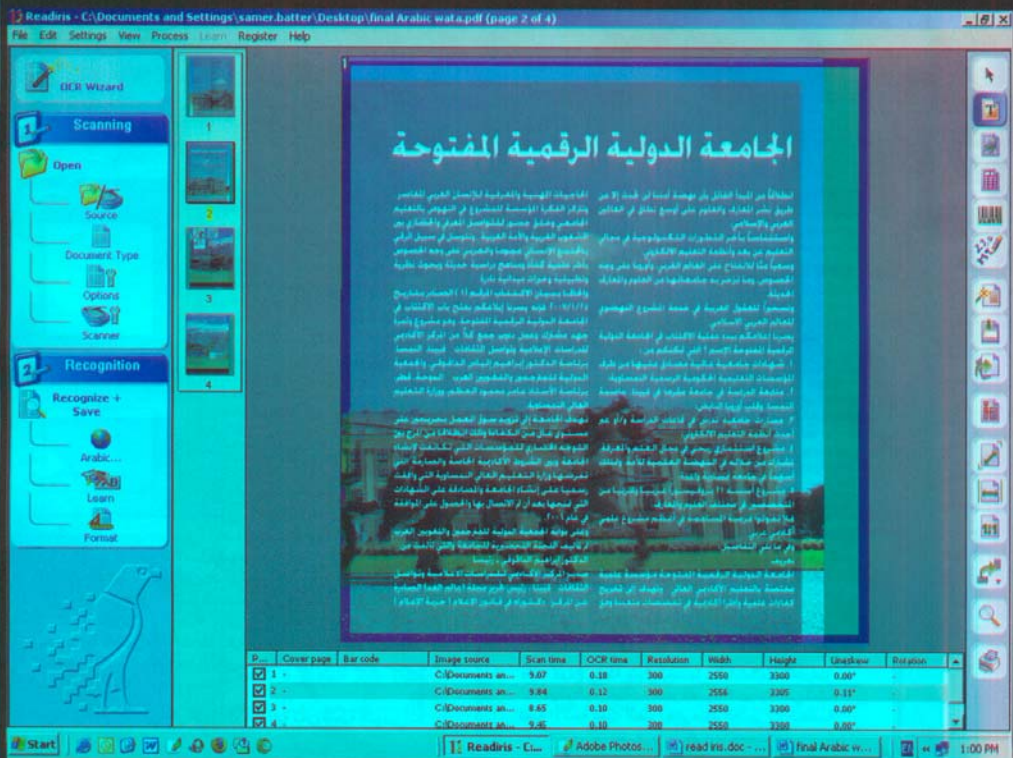
الأمر الذي يجب مراعاته عند العمل مع ريد أيريس هو تحديد لغة المستند قبل البدء بعملية المسح، وكما في برنامج فيروس VEROUS لا يعطي ريد أيريس نتائج جيدة عند احتواء المستند على أكثر من لغة لأنه لا يدعم كما في برنامج صخر ضبط التعامل مع المستندات التي تحتوي على أكثر من لغة.

بالانتقال إلى النتائج التي حققها البرنامج في الاختبارات نجد أنه يأتي في المرتبة الثانية في الجودة بعد برنامج فيروس، فعند مسح مستند نظيف يحتوي على نص عربي كانت نسبة الدقة ٨٤٪ بينما كانت عند مسح مستند عربي متنسخ ٧٤٪، وعند مسح مستند إنكليزي نظيف كانت نسبة الدقة ٩٦.٧٪ وعند استخدام مستند إنكليزي متنسخ في الاختبار كانت نسبة الدقة ٨٩٪.

يتمتع برنامج ريدأيريس برو بواجهات جذابة تميزه عن المنافسين، ولكن هذه الواجهات لا تساعد كثيرا على فهم البرنامج بسرعة وإتقان طريقة استخدامه.

ولكن ما يميزه عن البرنامجين الآخرين المعروضين في هذه المقارنة هو تقديم معالج مساعدة يمكن المستخدم من المرور في جميع الخطوات اللازمة ابتداء من تحديد مصدر المستند أي من المسح الضوئي مباشرة أم من ملف صورة، ومرورا بتحديد الشكل العام للمستند أي يحتوي على نصوص فقط أو نصوص وصور أو غير ذلك، وانتهاء بتحديد اللغة التي يحتويها المستند وإجراء عملية المسح.

يقوم ريدأيريس عند استيراد الصورة بإجراء عملية المسح بشكل تلقائي وتحديد المناطق التي تحتوي على النص مع إمكانية ضبط هذه المناطق يدويًا.



يميز برنامج ريد أيريس الواجهة المحيطة واحتواءه على معالجات للمساعدة على استخدامه، بالإضافة إلى التفاصيل الدقيقة التي يقدمها حول عملية المسح.

برامج التعرف على الأحرف العربية

اختبار جماعي

NovoDynamics

النسخة Verus Standard
السعر ١٢٩٩ دولار
الاتصال +٩٧٤٢٩٧٣٧٧٥
الموقع www.novodynamics.com

توحي واجهة برنامج فيروس Verus بصعوبة استخدامه ويحمل هذا الشعور شيئاً من الصواب، فهو ليس بتلك السهولة التي يتمتع بها برنامج صخر.

فور استيراد ملف الصورة التي تود التعرف على الأحرف التي تحتويها يبدأ البرنامج بعمليات التنظيف والتحليل والمعالجة واستخراج النص وكل ذلك يتم بشكل تلقائي، إذ يقوم البرنامج بالتعرف على المناطق التي تحتوي على النص وتحديد هذه المناطق، ويتبع بعد ذلك كما البرامج المنافسة إمكانية تحديد منطقة معينة واستخراج النص من هذه المنطقة، وعلى الرغم من الزمن الطويل نسبياً الذي يستغرقه فيروس لإجراء العمليات السابقة التعرف على النص المطلوب إلا أن نسب الدقة التي حققها تستحق أن نرفع القبة احتراماً لهذا البرنامج لما يقدمه من أداء.

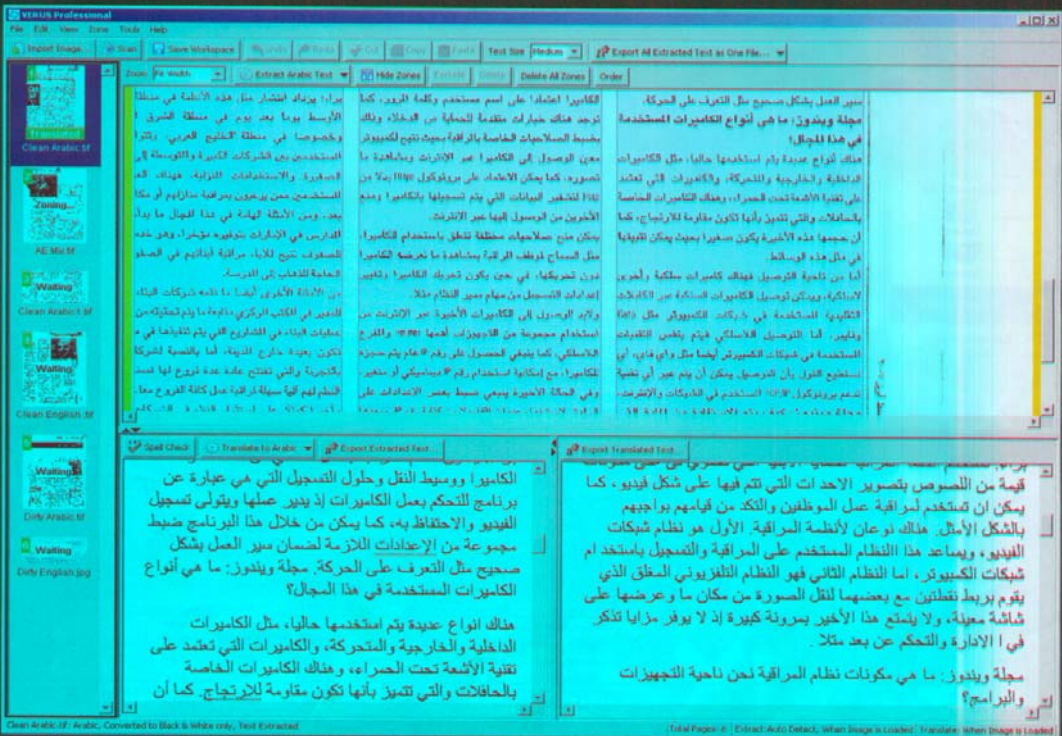
عند إجراء عملية المسح على مستند تخليف يحتوي على نص عربي كانت نسبة الدقة التي حققها البرنامج ٩٤.٥٪ وعند إجراء عملية المسح على مستند عربي ولكن

ليس بتلك النظافة كانت نسبة الدقة ٧١٪، أما بالنسبة للمستندات التي تحتوي على نص إنكليزي، فعند مسح مستند نظيف كانت نسبة الدقة ٩٧.٣٪ وبالنسبة للمستند الإنكليزي غير الخفيف كانت نسبة الدقة ٧٢٪، أما زمن عملية المسح فكان أكبر زمن بالمقارنة مع البرامج المنافسة، في حين يدعم برنامج صخر أكثر من عشر لغات يقتصر فيروس على خمسة لغات هي العربية والإنكليزية والفارسية والدرية والبشتو بالإضافة إلى إمكانية تخصيص حجم الخط بين الصغير والكبير والمتوسط لمساعدة البرنامج على التعرف على الأحرف بالشكل الصحيح، ولا بد ليعمل البرنامج بشكل سليم من أن تكون دقة الصورة ٣٠٠ نقطة في الإنش على الأقل.

يوفر برنامج فيروس إمكانية تنظيف المستند في حال احتوائه على بعض التشوهات قبل البدء بعملية المسح للتعرف على الأحرف، ولكن الناحية السلبية التي يمكن الحديث عنها هي الوقت الطويل الذي يستغرقه البرنامج في عملية المعالجة والتحليل، لذلك سرعان ما

ستنسى هذه الناحية عند رؤية النتائج الرائعة التي يحققها فيروس بعد عملية المسح. ومن النواحي السلبية التي لاحظناها على البرنامج أيضاً حدوث الكثير من الأخطاء عند احتواء المستند على أكثر من لغة، إذا لا يمكن ضبط البرنامج بحيث يبحث عن لغتين معاً كما في برنامج صخر، فعند اختبار التعرف على محارف العربية من القائمة اليمينة Extract Text>Arabic سيتعرف البرنامج على الأحرف العربية ولكن سيظهر الكثير من الأخطاء في التعرف على الأحرف الإنكليزية التي تتخلل النص والعكس صحيح.

بعد التعامل مع البرنامج لفترة وتجريب أنواع مختلفة من ملفات الصور النقطية والمسح لاحتفظ نفس المشكلة والتي تكمن في الوقت الطويل الذي تستغرقه معالجة الصور وتحليلها واستخراج النصوص منها، ولو عملت الشركة في المستقبل على تحسين سرعة عملية المعالجة لزادت قوة البرنامج ولأصبح الحل الذي تبحث عنه أغلب الشركات التي تتعامل مع المستندات بكثرة.



لا يقل برنامج فيروس بسهولة عن برنامج صخر وهذا مايتضح، ويظهر في الأسفل مربعي نص أحدهما هو النص الناتج عن المسح والآخر هو النص الصحيح.

اختبار جماعي

برامج التعرف على الأحرف العربية

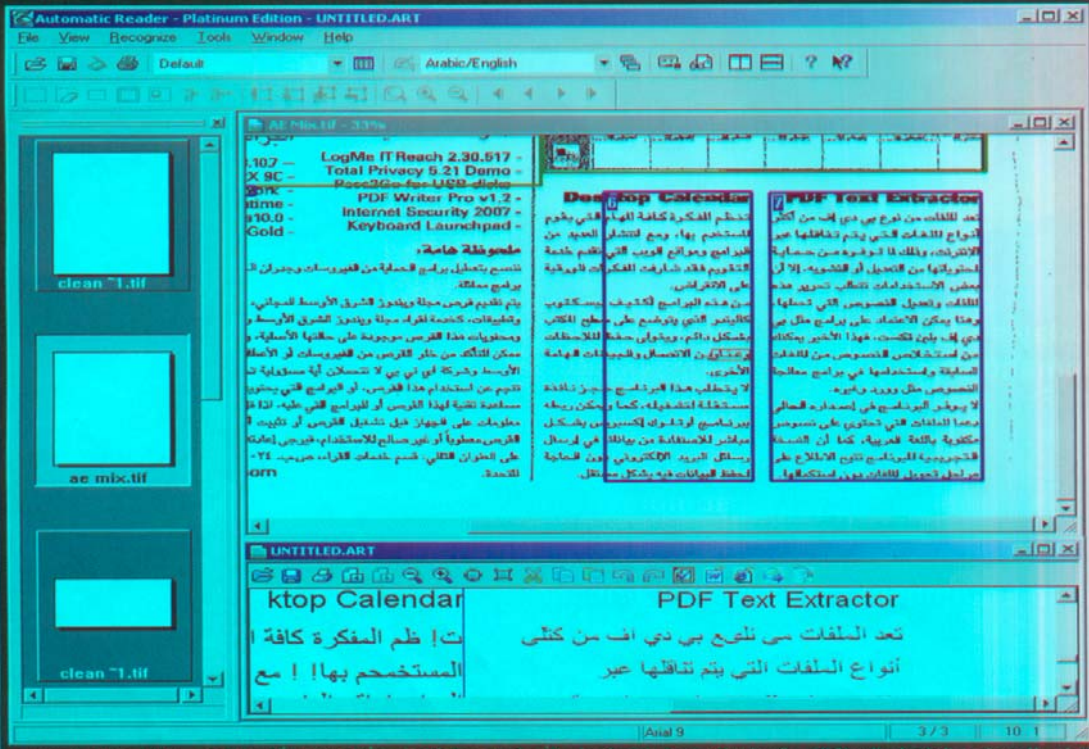
Sakhr

Automatic Reader 9.0
المنتج
السعر ١٤٠٠ دولار
الاتصال +٩٧١٢٧٦٢٨٨٨
الموقع www.sakhr.com

واستغرق ستة ثواني بينما استغرقت كل من عمليات المسح السابقة حوالي ثلثيتين. وقد قمنا بتجريب البرنامج باستخدام مستندات مختلطة تحتوي على العربية والإنكليزية وكانت النتيجة هي الحصول على النص الإنكليزي تقريبا بدون أخطاء، بينما كانت نسبة الدقة في اللغة العربية قريبة من ٥٠ بالمئة. بحسب البرنامج صخر أنه يدعم التعامل مع عدد كبير من اللغات فعدا عن العربية والإنكليزية يدعم البرنامج التعامل مع اللغة الفرنسية والفارسية وأوردو وبشتو والألمانية والفنلندية والتركية والسويدية والروسية والبرتغالية. يستخدم برنامج صخر محرك أومني Omni للتعرف على الأحرف، وتقدم صخر حزمة SDK التي يمكن استخدامها من قبل المطورين الآخرين في بناء التطبيقات التي تدعم التعرف على الأحرف، وهذه الحزمة متوفرة في النسخة البلاطينية من برنامج صخر والتي تباع بسعر ٤٠٠٠ دولار. وتوفر صخر نسخة أقل سعرا هي النسخة الذهبية التي تباع بسعر ١٤٠٠ دولار.

بشكل صحيح، لأن عدم ضبطه يؤدي إلى نتائج سيئة. بالانتقال إلى الاختبارات التي أجريناها على ملفات الاختبار حقق برنامج صخر أرقاما ممتازة جدا عند التعامل مع مستندات باللغة الإنكليزية ولكن هذه الأرقام كانت أقل عند التعامل مع ملفات اللغة العربية، فعند مسح مستند عربي نظيف (أي أن الورقة الأصلية غير متسخة) كانت نسبة الدقة ٥٧.٨٪ أي أن الكلمات الصحيحة الناتجة عن المسح كانت أكثر من الكلمات الخاطئة وكان معدل المسح ٢٧٧٧ حرف في الثانية، وعند إجراء عملية المسح على مستند عربي متسخ بعض الشيء، كانت نسبة الدقة ٤٨.٦٪ أي بعبارة أخرى كان عدد الكلمات الصحيحة يساوي تقريبا عدد الكلمات الخاطئة وكان معدل المسح ٢١٥٧ حرف في الثانية. أما بالنسبة للمستندات الإنكليزية فكانت نسبة الدقة فيها عند التعامل مع مستند نظيف ٩٨.٥٪ ومعدل المسح ٢٨٤٦ حرف في الثانية أما عند التعامل مع مستند إنكليزي متسخ فكانت نسبة الدقة ٩٧.٦٪ ومعدل المسح ٢٢٧ حرف في الثانية

يتميز برنامج صخر ببساطة واجهة الاستخدام فلم نستغرق وقتا طويلا لنألف استخدامه. يحتاج البرنامج ليعمل بالشكل الصحيح أن تكون الصورة المراد استخلاص النص منها بدقة ٢٠٠ نقطة في الإنش حصرا ولا يعمل البرنامج مع الصور ذات الدقة الأقل. تحتوي واجهة البرنامج على قائمتي اختيار لتعيين الدور الأساسي في عمل البرنامج بشكل صحيح، الأولى هي لتحديد نوع الصورة المستخدمة، أي هل هي ملونة أم بلونين فقط، هل تحتوي على جداول أم هي عبارة عن نص فقط. والثانية والأهم هي لاختيار اللغة التي ترغب في التعامل معها أي لغة المستند الذي تتعامل معه، ويوفر البرنامج مجموعة واسعة من الخيارات منها اللغة العربية واللغة الإنكليزية ويمكن اختيار خليط بين لغتين عند التعامل مع المستندات التي تحتوي على أكثر من لغة كاختيار العربية والإنكليزية معا أو فارسي وإنكليزي أو عربي وفرنسي وخيارات أخرى كثيرة. أهم ما يجب ضبطه قبل استخدام برنامج صخر هو ضبط هذا الخيار



لا داعي للحدوث عن الواجهات السهلة التي يقدمها برنامج صخر للتعرف على الأحرف العربية، فالصورة أعلاه تحكي ألف كلمة عن سهولة استخدام البرنامج.