تحسين مرحلة "بعد المعالجة" في نظام التعرف الضوئي الآلي على الكتابة العربية

حسني المحتسب وحمزة لقمان

مقدمة

بالرغم من تطور التقنيات والطرق المستخدمة في نظام التعرف الضوئي الآلي على الكتابة العربية (OCR) إلا أنه مازال هناك بعض الأخطاء التي تظهر في النص الذي يتم التعرف عليه. وتقتصر الطرق التقليدية في مرحلة بعد المعالجة بتحديد صحة الكلمة من عدمها باستخدام المعجم دون إجراء أي تصحيح للكلمات غير الصحيحة. تستعرض هذه الورقة عدة طرق لمعالجة الأخطاء الناتجة عن التعرف الضوئي للنص العربي، التي بدورها تساعد في تقليل نسبة الخطأ في نظام التعرف الضوئي الآلي على الكتابة العربية.

يتم في مرحلة "بعد المعالجة" (Post-Processing) تصحيح النص الذي تم التعرف عليه بواسطة نظام التعرف الضوئي الآلي على الكتابة العربية. حيث يحتوي النص الذي تم التعرف عليه (الناتج من التعرف الضوئي للأحرف) غالباً على عدة أخطاء نتيجة التعرف الخاطئ للأحرف أو نتيجة إدراج أو حذف أحرف ليست من أصل الكلمة.

وقد خلصت بعض الدراسات[١] إلى إمكانية تقليل نسبة الأخطاء في نظام التعرف الآلي على الكتابة العربية المطبوعة إلى النصف (٥٠٪) وذلك من خلال إجراء بعض العمليات على الكلمات غير الصحيحة بهدف إرجاعها إلى كلمات صحيحة. ويلاحظ الباحث في هذا المجال قلّة التقنيات المطورة لمعالجة الأخطاء الناتجة عن أنظمة التعرف الضوئي للكتابة العربية. اقترح يوسف وعلوني [٢] طريقة لمعالجة الأخطاء الناتجة عن أنظمة التعرف الضوئي للكتابة العربية باستخدام المدقق الإملائي الخاص بشركة جوجل على الإنترنت لمعالجة الكلمات غير الصحيحة. أما بالنسبة للغات الأخرى فقد عمل الباحثون على تطوير عدد من الطرق لمعالجة الأخطاء الناتجة عن أنظمة التعرف الضوئي للكتابة، ويمكن الرجوع لبعضها بمراجعة البحوث المثبتة في قائمة المراجع [٣-١٢].

ويمكن تقسيم الأخطاء التي يمكن أن تظهر في الكلمة الناتجة عن التعرف الضوئي الآلي على الكتابة العربية إلى نوعين: أحادي ومتعدد. فالخطأ الأحادي هو خطأ ينتج عن استبدال أو إدراج أو حذف حرف واحد من الكلمة الأصلية بشكل غير صحيح. والأخطاء المتعددة هي الأخطاء التي تنتج عن وجود أكثر من خطأ في الكلمة الواحدة كاستبدال أكثر من حرف في الكلمة الأصلية بأحرف أخرى غير صحيحة.

ومن هنا يأتي دور النظام المعروض في هذا البحث في معالجة الأخطاء التي تنتج عن عملية التعرف الضوئي الآلي على الكتابة العربية. حيث يتم معالجة أخطاء الكلمات ابتداءً من معالجة الأخطاء الأحادية ثم الأخطاء المتعددة. ويتم معالجة الخطأ الأحادي في الكلمة من خلال إجراء عمليات استبدال أو إدراج أو حذف لأحرف الكلمة وذلك للوصول إلى كلمة صحيحة. وفي حالة تعذر التصحيح باستخدام المعالجة الأحادية (معالجة حرف واحد في الكلمة) يتم التحول إلى المعالجة المتعددة التي تبنى على فرضية وجود أكثر من خطأ في الكلمة الواحدة. وفي مرحلة المعالجة المتعددة يتم معالجة الخطأ من خلال إجراء استبدال أو إدراج أو حذف أو استبدال وحذف أو استبدال وإدراج أو حذف وإدراج لأكثر من حرف من الكلمة حتى نصل إلى كلمة صحيحة.

يعرض هذا البحث طرق معالجة كل أنواع الأخطاء الناتجة عن مرحلة التعرف الضوئي الآلي على الكتابة العربية التي من خلالها يمكن تقليل نسبة الخطأ إلى أكثر من ٥٧٪ وفقاً للنتائج التي توصلنا إليها.

منهج الدراسة

عند الرجوع إلى أصل الأخطاء، كما سيتم توضيحه في الجزء التالي، فإنه بالإمكان معالجة الأخطاء الناتجة عن التعرف الضوئي الآلي على الكتابة العربية من خلال طريقة معاكسة بحيث يتم إرجاع الكلمة التي تحتوي أخطاء إلى كلمات صحيحة كتلك التي في النص الذي تم التعرف عليه. ولإنجاز هذا الهدف يجب في البداية معرفة الخطأ ومعالجته وفقاً لنوع الخطأ الذي في الكلمة.

أنواع الأخطاء وطرق معالجتها

تختلف طريقة معالجة الكلمات غير الصحيحة بناءً على نوع الخطأ الذي في الكلمة، فهناك أخطاء أحادية وأخطاء مزدوجة. نناقش في الأجزاء التالية من البحث كل نوع من الأخطاء المعالجة مع ذكر طريقة معالجتها.

أولاً: الأخطاء الأحادية وطرق معالجتها

وهي الأخطاء التي تنتج عن افتراض وجود خطأ واحد فقط في الكلمة، وهي كالتالي: -

أخطاء الاستبدال (Substitution Errors)

ينتج هذا الخطأ عند التعرف على حرف معين بشكل غير صحيح بواسطة نظام التعرف الضوئي الآلي على الكتابة العربية. فمثلاً تبديل حرف (ف) بحرف (غ) في كلمة (فساد) ليتم التعرف عليها (غساد)، حيث نلاحظ أن حرف الفاء تم التعرف عليه بشكل غير صحيح مما نتج عنه كلمة غير صحيحة. ويعتبر هذا النوع من الأخطاء أكثر أنواع الأخطاء شيوعاً في نظام التعرف الضوئي الآلي على الكتابة العربية.

معالجة أخطاء الاستبدال (Substitution Errors Correction)

يتم معالجة أخطاء التبديل من خلال إرجاع كل حرف إلى أصله المنقلب منه. حيث يتم العمل على كل حرف من حروف الكلمة حتى نصل لكلمة صحيحة أو نستنفد كل حروف الكلمة فننتقل إلى معالجة مختلفة. ولكي نتعرف على أقرب حرف يمكن أن يتبدل به الحرف المعنى نستعين بمصفوفة الالتباس (Confusion Matrix) التي تحوي احتمالات تبديل كل حرف بالحروف الأخرى. حيث يتم تبديل الحرف بالحرف المتبدل ثم نبحث عن الكلمة الجديدة في القاموس. وتستمر هذه العملية إلى حين الوصول إلى كلمة صحيحة أو انتهاء تجربة التبديل دون أي نتيجة والذي بناءً عليه ننتقل إلى المعالجة بالإدراج.

أخطاء الإدراج (Insertion Errors)

هي الأخطاء التي تنتج عن إدراج حرف أثناء التعرف على الكلمة ليس من أصل الكلمة. فمثلاً إدراج حرف الباء على كلمة (استيقظ) لتصبح الكلمة بعد التعرف عليه (استيبقظ) والتي تعتبر كلمة غير صحيحة.

معالجة أخطاء الإدراج (Insertion Errors Correction)

لمعالجة هذا الخطأ الناتج عن الإدراج الخاطئ لحرف على الكلمة، يتم حذف الحرف الذي لديه أعلى احتمالية إدراج على الكلمة وذلك بالاستعانة بمصفوفة الالتباس التي تحوي احتماليات إدراج كل حرف، ثم التأكد من صحة الكلمة الناتجة فإذا وجد أن الكلمة ما زالت غير صحيحة فيتم الانتقال إلى الحرف الآخر (الذي لديه احتمالية إدراج أقل) وحذفه والتأكد من الكلمة الجديدة، وتستمر عملية تجربة حذف حرف من الكلمة حتى نصل إلى كلمة صحيحة أو استنفاد كافة الحروف التي يمكن أن تدخل على الكلمة دون الوصول إلى كلمة صحيحة، وفي هذه الحالة (عدم العثور على كلمة صحيحة) يتم الانتقال إلى المعالجة بالحذف.

أخطاء الحذف (Deletion Errors)

هي الأخطاء التي تنتج عند حذف أحد حروف الكلمة أثناء التعرف عليها فتنتج كلمة غير صحيحة. فمثلاً حذف حرف التاء في كلمة (استنفاد) ليتم التعرف عليها (اسنفاذ) والتي تعتبر غير صحيحة.

معالجة أخطاء الحذف (Deletion Errors Correction)

يتم معالجة هذا النوع من الأخطاء بالرجوع إلى مصفوفة الالتباس التي تحوي احتمالات الحروف التي يمكن أن تحذف من الكلمة، حيث يتم ترتيبها تنازلياً حسب احتمالية الحذف، بعد ذلك يتم إدراج الحرف ذي الاحتمالية العالية على جميع مواقع الكلمة، وتستمر عملية الإدراج إلى حين الوصول إلى كلمة صحيحة أو انتهاء الحروف التي يمكن أن تحذف، وفي هذه الحالة يتم الانتقال إلى المرحلة الثانية من المعالجة والتي تفترض وجود أكثر من خطأ واحد في الكلمة.

ثانياً: الأخطاء المتعددة وطرق معالجتها

الأخطاء المتعددة هي الأخطاء التي تنتج عن وجود أكثر من خطأ في الكلمة الواحدة. فيتم معالجتها بناءً على نوع الخطأ الذي يمكن أن يوجد في الكلمة. وهناك عدة أنواع من الأخطاء المتعددة نذكر أهما في الأقسام التالية.

أخطاء استبدال متعدد

هي الأخطاء التي تنتج عن استبدال حرفين من الكلمة بحرفين غير الحروف الأصلية التي في الكلمة، فمثلًا كلمة (تكريم) يتم التعرف عليها (نكربم) حيث نلاحظ أنه تم استبدال حرف التاء بالنون وحرف الياء بالباء والذي نتج عن كلمة غير صحيحة.

معالجة أخطاء استبدال حرفين

يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تستبدل في الكلمة الواحدة، حيث يتم ترتيب هذه الحروف بناءً على احتماليات تبدلها في الكلمة الواحدة بشكل تنازلي، ثم يتم إجراء الاستبدال لهذه الحروف حتى الوصول إلى كلمة صحيحة. وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم الانتقال إلى النوع الآخر من المعالجة والتي تبنى على أساس وجود استبدال وإدراج في نفس الوقت.

أخطاء استبدال وإدراج

هي الأخطاء التي تنتج عن استبدال حرف من الكلمة بحرف غير الحرف الأصلي وإدراج حرف جديد ليس في أصل الكلمة، فمثلاً كلمة (أودية) يتم التعرف عليها (اردبة) حيث نلاحظ أنه تم استبدال حرف الواو بالراء وحرف الياء بالباء والذي نتج عن كلمة غير صحيحة.

معالجة أخطاء استبدال وإدراج

يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تستبدل وتدخل في الكلمة الواحدة بشكل غير صحيح، حيث يتم ترتيب هذه الحروف بناءً على احتماليات تبدلها وإدراجها في الكلمة الواحدة بشكل تنازلي، ثم يتم إجراء الاستبدال والحذف على جميع مواقع الحروف في الكلمة حتى الوصول إلى كلمة صحيحة، وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم الانتقال إلى النوع الآخر من المعالجة والتي تبنى على أساس وجود استبدال وحذف في نفس الوقت.

أخطاء استبدال وحذف

هي الأخطاء التي تنتج عن استبدال حرف من الكلمة بحرف غير الحرف الأصلي وحذف حرف من أصل الكلمة، فمثلًا كلمة (سيارة) يتم التعرف عليها (سبار) حيث نلاحظ أنه تم استبدال حرف الياء بالباء وتم حذف حرف (ة) والذي نتج عنه كلمة غير صحيحة.

معالجة أخطاء استبدال وحذف

يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تستبدل وتحذف من الكلمة الواحدة بشكل غير صحيح، حيث يتم ترتيب هذه الحروف بناءً على احتماليات تبدلها وحذفها في الكلمة الواحدة بشكل تنازلي، ثم يتم إجراء الاستبدال والإدراج على جميع مواقع الحروف في الكلمة حتى الوصول إلى كلمة صحيحة. وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم الانتقال إلى النوع الآخر من المعالجة والتي تبنى على أساس وجود إدراج وحذف في نفس الوقت.

أخطاء إدراج وحذف

هي الأخطاء التي تنتج عن إدراج حرف زائد على الكلمة وحذف حرف أخر من أصل الكلمة، فمثلًا كلمة (خيول) يتم التعرف عليها (خوال) حيث نلاحظ أنه تم حذف حرف الياء وتم إدراج حرف الألف والذي نتج عنه كلمة غير صحيحة.

معالجة أخطاء إدراج وحذف

يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تدرج والحروف التي يمكن أن تحذف بشكل غير صحيح، حيث يتم ترتيب هذه الحروف بناءً على احتماليات إدراجها وحذفها بشكل تنازلي. ثم يتم إجراء حذف وإدراج على جميع مواقع الحروف في الكلمة حتى الوصول إلى كلمة صحيحة. وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم الانتقال إلى النوع الآخر من المعالجة والتي تفترض وجود إدراج مكرر.

أخطاء إدراج متعدد

هي الأخطاء التي تنتج عن إدراج أكثر من حرف في الكلمة، فمثلا كلمة (مستقبل) يتم التعرف عليها (مستتقببل) حيث نلاحظ أن حرفي التاء والباء مدخله على الكلمة بشكل غير صحيح.

معالجة أخطاء إدراج وإدراج

يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تدخل على الكلمة بشكل غير صحيح. حيث يتم ترتيب هذه الحروف بناءً على احتماليات إدراجها بشكل تنازلي، ثم يتم إجراء حذف لهذه الحروف من الكلمة حتى الوصول إلى كلمة صحيحة. وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم الانتقال إلى النوع الآخر من المعالجة والتي تبنى على أساس حذف أكثر من حرف من الكلمة في نفس الوقت.

أخطاء حذف متعدد

هي الأخطاء التي تنتج عن حذف أكثر من حرف على الكلمة، فمثلا كلمة (ينابيع) يتم التعرف عليها (يابع) حيث نلاحظ أن حرفي

النون والباء تم حذفهما من الكلمة بشكل غير صحيح.

معالجة أخطاء حذف وحذف

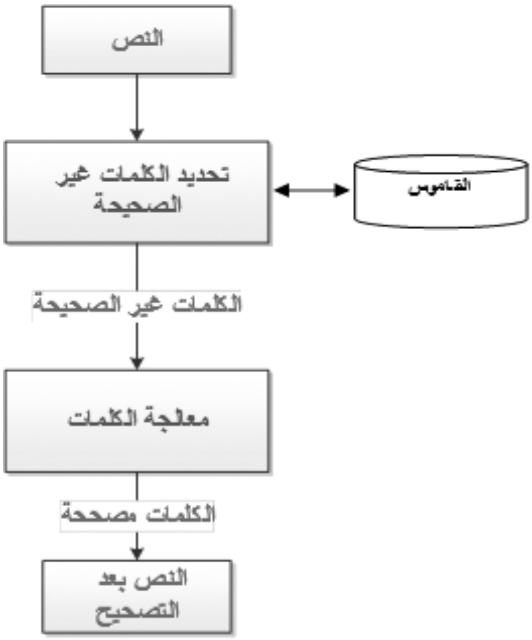
يتم معالجة هذا النوع من الأخطاء من خلال الاستعانة بمصفوفة الالتباس التي يمكن من خلالها معرفة الحروف التي يمكن أن تحذف من الكلمة بشكل غير صحيح. حيث يتم ترتيب هذه الحروف بناءً على احتماليات حذفها بشكل تنازلي، ثم يتم إجراء إدراج لهذه الحروف على جميع مواقع الكلمة حتى الوصول إلى كلمة صحيحة. وتنتهي هذه العملية عند الوصول إلى كلمة صحيحة أو الانتهاء من الحروف دون الوصول إلى كلمة صحيحة، وفي هذه الحالة يتم تقرير عدم أمكانية معالجة الكلمة.

التصميم العام للنظام

يبين المخطط رقم ١ التصميم العام للنظام. ويقوم النظام بإجراء العمليات التالية: –

- قراءات النص المتعرف عليه بواسطة نظام التعرف الضوئي الآلي على الكتابة العربية.
- التعرف على الكلمات غير الصحيحة
  - معالجة الكلمات غير الصحيحة معالجة أحادية
  - معالجة الكلمات غير الصحيحة معالجة متعددة
- عرض النص بعد التصحيح
- عرض الإحصائيات

وللفائدة نعرض مخطط رقم ٢ الذي يشير إلى مراحل التعرف الضوئي على الكتابة.



خوارزميات النظام

مرحلة بعد المعالجة

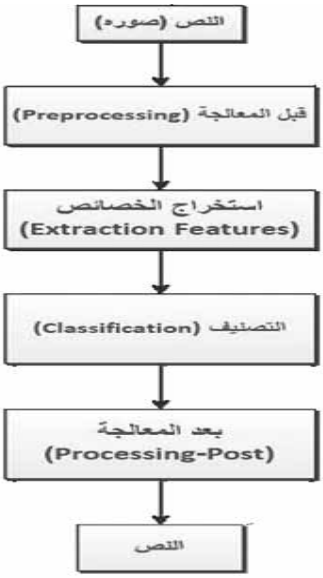
هي أخر مراحل التعرف الضوئي الآلي على الكتابة العربية التي من خلالها يتم تصحيح الكلمات التي تم التعرف عليها.

إعداد مصفوفة الالتباس

مصفوفة الالتباس هي مصفوفة يتم من خلالها معرفة إحصاءات عدد مرات استبدال أو إدراج أو حذف كل حرف الحروف العربية. ويتم إنشاء هذه المصفوفة في مرحلة التدريب والاختبارات المتكررة للوصول إلى أدق نتيجة ممكنة. وقد استفدنا جدا من مئات مصفوفات الالتباس الناتجة عن بحث المحتسب [١].

معالجة الكلمات غير الصحيحة (الخوارزمية الرئيسية للنظام)

يوضح مخطط رقم ٣ الخطوات العامة لمعالجة الكلمات غير الصحيحة.



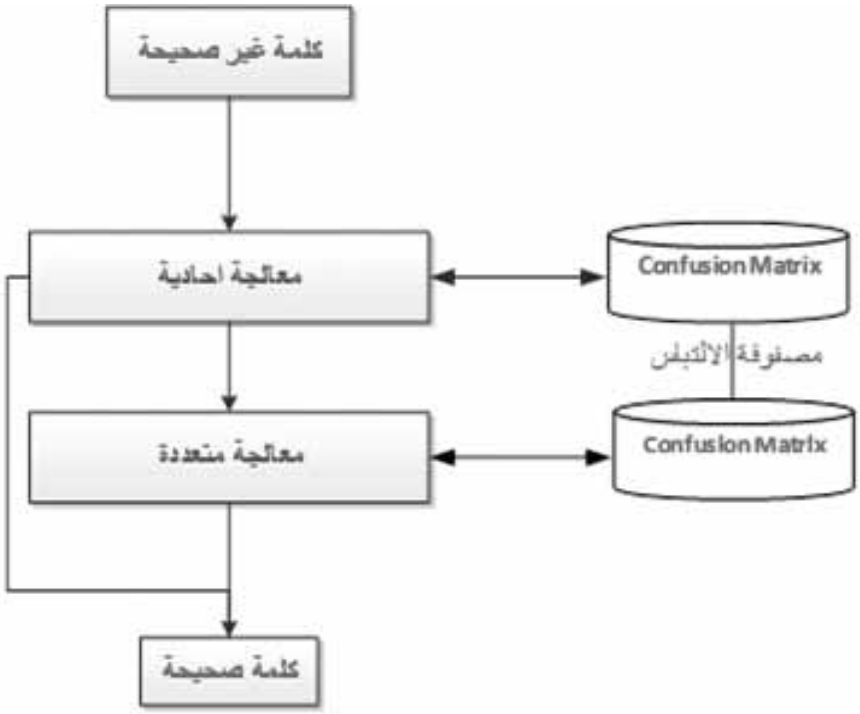
المعالجة الأحادية (خوارزمية المعالجة الأحادية)

يوضح مخطط رقم ٤ خطوات معالجة الأخطاء الأحادية والتي تشمل المعالجة بالتبديل ومعالجة أخطاء الإدراج ومعالجة أخطاء الحذف. ونعطي مثالا على هذه الخوارزميات في مخطط رقم ٥ الذي يظهر تسلسل سير عمليات معالجة أخطاء التبديل الأحادية.

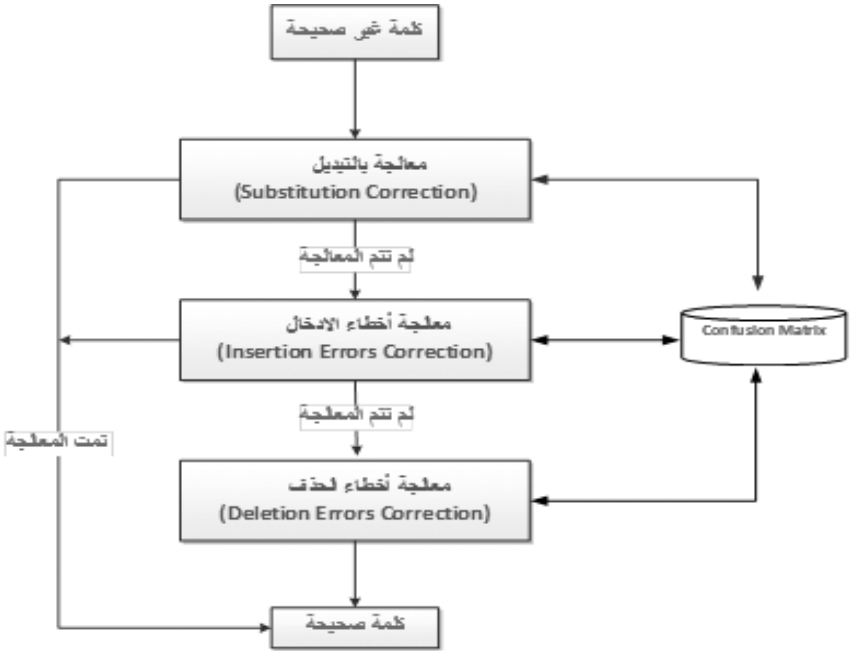
المعالجة المتعددة

بعد المرور على المعالجة الأحادية وعدم التمكن من تصحيح الكلمة يتم الانتقال إلى المعالجة المتعددة التي تفترض وجود أكثر من خطأ في الكلمة (عمليا خطأين). ويبين مخطط رقم ٦ الخوارزمية المستخدمة في المعالجة المتعددة، كما يبين مخطط رقم ٧ خوارزمية معالجة تبديل متكرر كمثال على الأخطاء المتعددة والتي تشمل التبديل المتكرر، والاستبدال مع الإدراج، والاستبدال مع الحذف، والإدراج مع الحذف والإدراج المتكرر، والحذف المتكرر.

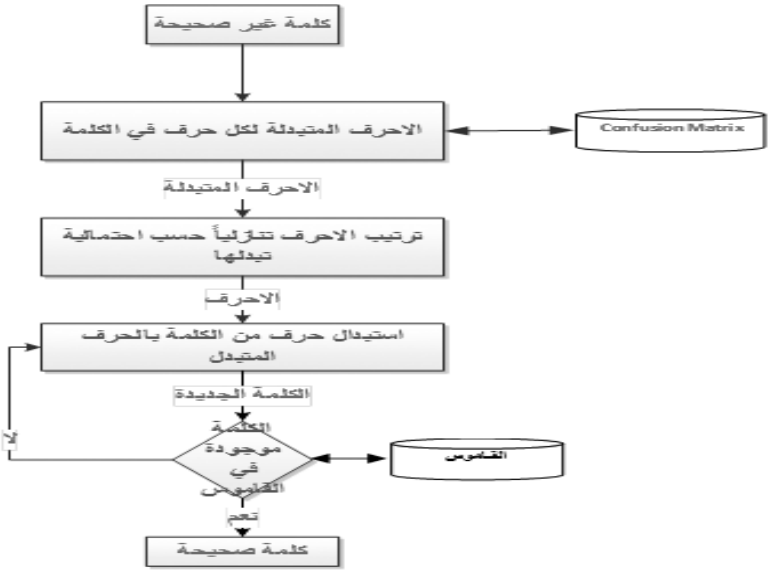
مخطط رقم ٣ خوارزمية معالجة الكلمات غير الصحيحة



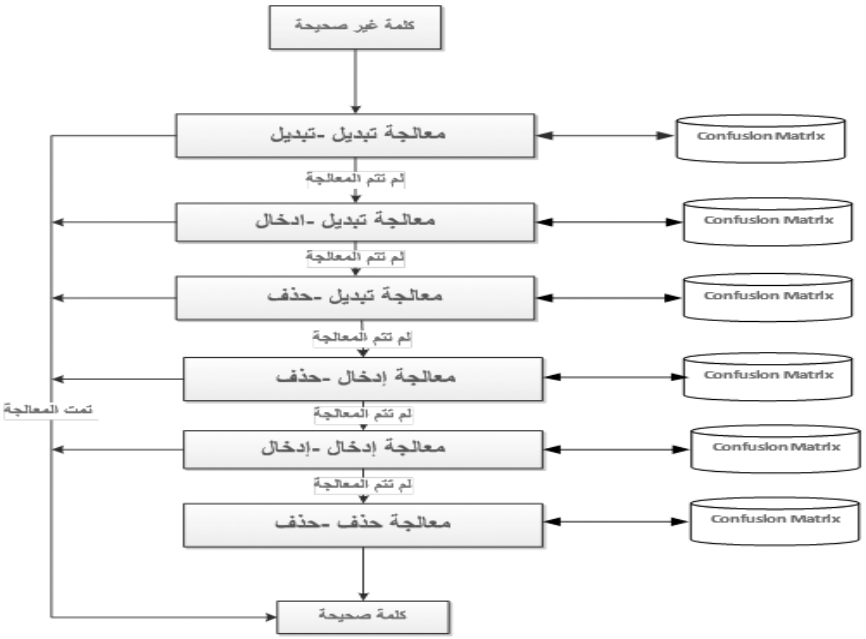
مخطط رقم ٤ خوارزمية المعالجة الأحادية



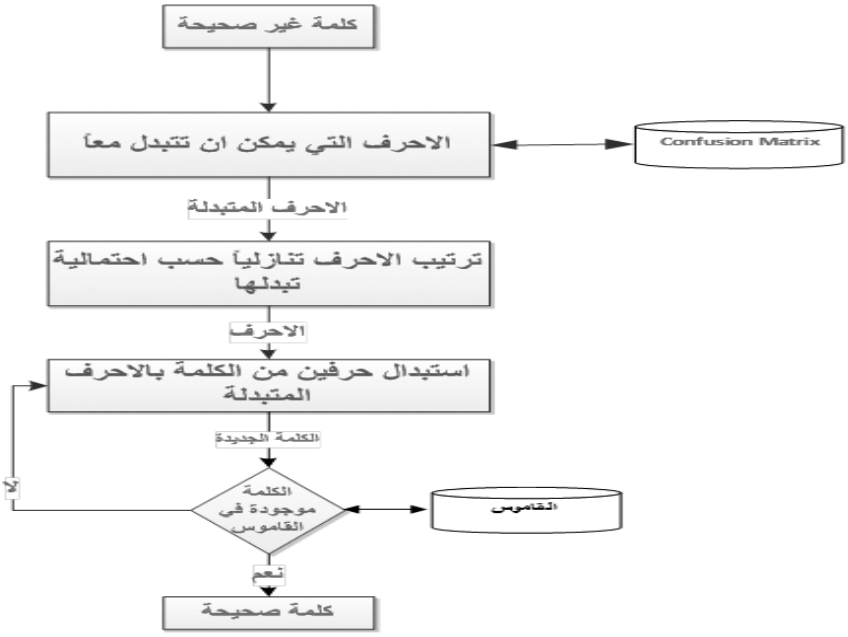
مخطط رقم ٥ خوارزمية معالجة استبدال حرف



مخطط رقم ٦ خوارزمية معالجة الأخطاء المتعددة



مخطط رقم ٧ خوارزمية معالجة أخطاء تبديل متكرر



النتائج

من خلال المعالجة الأحادية والمتعددة تم الوصول إلى نتائج جيده ساهمت في تقليل نسبة الخطأ بمقدار لا يقل ٥٠٪. وهذا يساهم في رفع نسبة التعرف في نظام التعرف الضوئي للكتابة. فبعد معالجة نصوص تحتوي على كلمات تم التعرف عليها بشكل خاطئ تم معالجة ٥٧,٦٩٪ من إجمالي هذه الأخطاء. وبالنظر في الكلمات التي تم معالجتها نجد أن نسبة الأخطاء الأحادية مثلت ٨٦,٦٪ وباقي الأخطاء عبارة عن أخطاء متعددة (استبدال وحذف، إدراج واستبدال،...الخ). ومن الكلمات المعالجة تم معالجة ٤٠٪ من الكلمات عن طريق المعالجة بالإدراج و٤٦٪ عن طريق المعالجة بالحذف، في حين تم معالجة ١٣,٤٪ عن طريق المعالجة المزدوجة (تبديل وإدراج).

التوصيات

تعتبر النتائج التي وصلنا إليها جيدة جداً، ويمكن زيادة نسبة التعرف (تقليل الخطأ) من خلال استخدام المحلل الصرفي الذي سوف يساعد على التعرف على الكلمات بعد إرجاع الكلمة إلى جذرها وإزالة السوابق واللواحق من الكلمة. كذلك يمكن زيادة دقة التصحيح من خلال ربط الكلمات التي تم التعرف عليها بموضوع النص الذي تم التعرف عليه.

الخاتمة

مرحلة بعد المعالجة هي المرحلة التي تهتم بتصحيح النص الذي تم التعرف عليه بنظام التعرف الضوئي الآلي على الكتابة العربية. والأخطاء التي تنتج في النص المتعرف عليه تكون أخطاء أحادية أو أخطاء متعددة. الأخطاء الأحادية هي الأخطاء الناتجة عن استبدال أو إدراج أو حذف حرف من الكلمة، بينما الأخطاء المتعددة هي الأخطاء الناتجة عن وجود أكثر من خطأ في الكلمة. بعد إجراء معالجة للأخطاء الأحادية والأخطاء المتعددة تم تنقيص نسبة الخطأ بأكثر من ٥٠٪ الأمر الذي يساهم بشكل كبير في زيادة دقة أنظمة التعرف الضوئي على الكتابة.

المراجع

[١]. Husni Al-Muhtaseb. "Arabic Text Recognition of Printed Manuscripts". PhD Thesis. School of Computing. Informatics & Media. University of Bradford. UK. ٢٠١٠. (Supervisor: Dr. Rami Qahwaji).

[٢]. Bassil. Youssef. and Mohammad Alwani. OCR post-processing error correction algorithm using google online spelling suggestion. arXiv preprint arXiv:.(٢٠١٢) ١٢٠٤,٠١٩١

[٣]. Levenshtein. V.I.. Binary codes capable of correcting deletions. insertions. and reversals. Cybernetics and Control Theory. ١٩٦٦) ،٧١٠-٧٠٧ ،(٨)١٠).

[٤]. I. Guyon. F. Pereira. Design of a Linguistic Postprocessor Using Variable Memory Length Markov Models. in Proc. ٣rd Int. Conf. Document Analysis and Recognition. pp.٤٥٧-٤٥٤. Montreal. Canada. (١٩٩٥).

[٥]. Hisao Niwa. Kazuhiro Kayashima. Yasuham Shimeki. Postprocessing for Character Recognition Using Keyword Information. IAPR Workshop on Machine Vision Applications. Dec. ٩-٧. Tokyo. (١٩٩٢).

[٧]. Lon-Mu Liu. Yair M. Babad. Wei Sun. Ki-Kan Chan. Adaptive Post-Processing of OCR Text via Knowledge Acquisition. Proceedings of the ١٩th annual conference on Computer Science. (١٩٩١).

[٨]. Kazem Taghva. Julie Borsack. Allen Condit. Results of Applying Probabilistic IR to OCR Text. Proceedings of the seventeenth annual international ACMSIGIR conference on Research and development in information retrieval. (١٩٩٤).

[٩]. Michael L. Wick. Michael G. Ross. Erik G. Learned-Miller. Context-Sensitive Error Correction: Using Topic Models to Improve OCR. In the proceedings of the ٩th International Conference on Document Extraction and Analysis. page(s): ١١٧٢ - ١١٦٨. ISSN: ٢٠٠٧) ،٥٣٦٣-١٥٢٠).

[١٠]. K. Kise. T. Shiraishi. S. Takamatsu. K. Fukunaga. A method of Post-Processing for Character Recognition based on Syntactic and Semantic Analysis of Sentences. Journal of Systems and computers. Japan. ISSN ١٦٦٦-٠٨٨٢. CODEN SCJAEP. vol. ٢٧. no٩. pp. ١٩٩٦) ١٠٧-٩٤).

[١١]. J. Hull. Incorporating Language Syntax in Visual Text Recognition with a Statistical Model. IEEE Transactions on Pattern Analysis and Machine Intelligence. ١٩٩٦) ،(١٢)١٨).

[١٢]. Wei Sun. Lon-Mu Liu. Weining Zhang. John Craig Comfort. Intelligent OCR Processing. Journal of the American Society for Information Science. volume ٤٣. issue ٦. pages ١٩٩٢) ،٤٣١-٤٢٢).