التدقيق الإملائي للكلمات العربية المطبوعة باستخدام التحليل الإحصائي

د. غيداء عبد العزيز الطالب ماهر أحمد أنور النعمة

قسم علوم الحاسبات

كلية علوم الحاسبات والرياضيات

جامعة الموصل

Email: mahernima@yahoo.com Email:ghaydatalibz@yahoo.com

رافدة حميد الطائي

قسم الحاسبات

كلية التربية

جامعة الموصل

Email: rafida_hd@yahoo.com

الاستلام 2008 / 05 / 19

القبول 2009 / 02 / 16

ABSTRACT

In this research we propose a statistical method and morpho-lexical analysis for correcting Arabic words as a post processor for Arabic words output from OCR systems. Dictionaries of words were built for the comparison to the attached word.

The present research uses multiple knowledge sources and basing on the Arabic language properties, statistical method, morpho-lexical analysis and dictionary look-up for error detection and correction. Correction of errors in this research depends on the type of possible error, which can be: transposing two adjacent letters, rejection, replacing an incorrect letter, inserting a missing letter, substitution errors, which are most frequently committed by the OCR systems.

المستخلص

يقترح هذا البحث طريقة إحصائية وتحليل مورفولوجي لتصحيح الكلمات العربية الخاطئة بوصفها وسيلة معالجة نهائية للكلمات العربية الناتجة من أنظمة التمييز الضوئي OCR. يلحق بالبحث قواميس بالكلمات بهدف مقارنة الكلمة تحت المعالجة مع مطابقتها في القاموس.

يستخدم هذا البحث عدد من مصادر المعرفة ويعتمد خصائص اللغة العربية، طريقة إحصائية، تحليل مورفولوجي، فضلاً عن البحث في القواميس لتحديد موقع الخطأ وتصحيحه. تتوقف عملية تصحيح الأخطاء على نوع الخطأ المحتمل، والذي يشمل القلب المكاني، وإبدال الحروف، وحذفها، وإضافتها الممكنة، وذلك باستبدال الحروف المنقلبة بتلك التي تشابهها، وتعويض الحروف المحذوفة بتلك المسموح بها داخل سياق الكلمة، وهي الأخطاء الأكثر حدوثاً في أنظمة التمييز الضوئي.

1. المقدمة

الكلمات هي الوحدات الأساسية للتخاطب بكل وسائل الاتصال بين الناس. ولكن الإشارات الكهربائية التي تنقل الكلمات قد تصل إلى الهدف بحالة غير سليمة، وأصبحت عملية اكتشاف الأخطاء الإملائية وتصحيحها من العمليات الأساسية في إدخال البيانات، وتنسيق الكلمات، وقراءة النصوص أليأً، ونظم تبادل البيانات، وتستخدم كذلك لزيادة المهارات الإملائية لدى الصغار من خلال التعليم الآلي للغات [1].

يمكن أن ينشأ الخطأ الإملائي داخل النصوص لعدة أسباب، منها:

أ- خطأ الآلة في تمييز بعض الحروف في نظم قراءة النصوص آليا.

ب- الخطأ البشري في إدخال البيانات.

ت- خطأ خلال نقل البيانات عبر شبكات الاتصال نتيجة لمصادر التشويش المختلفة أو خطأ في معدات الإرسال والاستقبال ذاتها.

فكان الهدف الرئيسي من هذا البحث هو رفع نسب التمييز للكلمات عندما يفشل نظام التمييز الضوئي للحروف (OCR- *Optical Character Recognition*) في ذلك.

تم استخدام أسلوبين رئيسين في تشخيص الكلمة الخاطئة إملائيأ وتصحيحها، وهما التحليل المورفولوجي والتحليل الإحصائي للنص العربي المطبوع، واستخدمت لغة C++ في برمجة الخوارزميات المتعلقة بالبحث.

2. الدراسات السابقة

بحوث عدة أنجزت لتصويب الأخطاء في أنظمة التمييز الضوئي للحروف. تلك البحوث اتخذت اتجاهين لتصويب تلك الأخطاء، احدها كان على مستوى الكلمة والآخر كان على مستوى العبارة. بعض البحوث المنجزة على مستوى الكلمة تضمنت: تحليل التذبذب في الحرف أو الكلمة بواسطة مجاميع الحروف (n-gram)، بحث في القاموس، احتمالات تحليل إحصائية؛ بينما استخدمت بعض البحوث المنجزة على مستوى العبارة تقانة n-gram للكلمات، القواعد، تصنيف الكلمات، السياق اللغوي، والسياق المرئي. وفيما يأتي بعض التقانات المستخدمة في بحوث تصويب الأخطاء.

افترض الباحثون في [3,2] أن يكون الخطأ في نظام OCR ناتج عن الضوضاء في قناة توصيل المعلومة. عالج هذان البحثان الأخطاء في النصوص من نوع إضافة حروف، حذف حروف، أو تبديل حروف. واستند البحثان إلى قاموس للبحث عن الكلمة الصحيحة.

بينما استخدم الباحثون في [4] التحليل الصرفي الحساس للسياق في تصويب أخطاء استرجاع المعلومات، وقارن الباحثون في تطبيقهم الصرف العربي الحساس وغير الحساس للسياق، فظهر لهم تحسن في النتائج باستخدام الصرف العربي الحساس للسياق.

نشر الباحثون في [5] مخطط شبكي محدود للتصحيح الآلي للأخطاء الإملائية باقتراح بدائل لتصويب الكلمة الخاطئة. تعرض هذه البدائل على المستفيد ليختار الأفضل منها حيث يتم توثيقها مع النص.

عرض نظام يتكون من جزئين في البحث [6] لتمييز الحروف وآخر لتمييز الكلمات. قُسم جزء تمييز الحروف إلى ثلاث أطوار: تمييز الحروف، تقطيع الكلمة إلى حروف, وتعريف الحروف. أما جزء تمييز الكلمات فانه استند إلى خوارزمية Viterbi.

عرض البحث [7] طريقة لتصويب الكلمات العربية الصادرة من نظام OCR باستخدام تحليل نوع (Morpho-Lexical) لتصويب خطأ رفض الكلمة أو خطأ إظهار كلمة غير صحيحة. استخدمت خصائص اللغة العربية في بناء قواعد الإنتاج Production Rules لنظام التصويب.

استخدم البحث [8] الوثائق العربية المتوفرة على الورق فقط والتي أمكن تحويلها بواسطة نظام OCR إلى صيغة مطبوعة يسهل استرجاعها والاستفادة منها. وقد استخدم الباحثون تقانات تصحيح مختلفة للكلمات فأوضحت النتائج بان تقليل الأخطاء في الكلمات يفيد في استرجاع وثائق أدق.

3. اللغة العربية

مجموعة الحروف العربية تضم 28 حرفاً أساسياً، خمسة عشر حرفا لها نقاط وثلاثة عشر حرفا بدون نقاط. النقاط فوق وتحت الحرف تلعب دوراً كبيراً في التمييز الآلي للحروف التي تختلف فيما بينها في عدد وموقع النقاط فقط (ب ت ث ي ن) هذه الحروف الخمسة تكتب بالطريقة نفسها في شكلها الوسطي والاختلاف الوحيد بينها هو بموقع وعدد النقاط. وهناك أربعة حروف قد ترافقها الهمزة 'ء'؛ هم الألف أ إ، الواو ؤ، الياء ئ، والكاف ك [9].

كما عالج الباحثون [10] في مجال التمييز الضوئي للحروف الصفات الأخرى للحروف العربية وهي:

1) الحروف العربية تكتب من اليمين إلى اليسار.
2) الحروف العربية (كتابة اليد أو المطبوعة) ليس لها عرض ثابت أو حجم ثابت.

3) بعض الحروف العربية لا تتصل مع الحروف التالية لهـا، وهـي (ا د ذ ر ز و) ممـا يضيف معالجة خاصة لهذه الحالة عند التعامل مع الكلمات العربية لأنها تقطع الكلمة إلى عدة مقاطع.

4) بعض الحروف العربية لها أشكال مختلفة حسب موقعها في الكلمة تصل إلـى أربعـة أشكال، كما مبين في (الجدول- 1) [11،9].

جدول (1): أشكال الحروف العربية حسب موقعها في الكلمة

الحروف مستقلة	الشكل الأولي	الشكل الوسطي	الشكل النهائي	أشكال أخرى
ا			ـا	ى ـى
ب	بـ	ـبـ	ـب	
ت	تـ	ـتـ	ـت	ة ـة
ث	ثـ	ـثـ	ـث	
ج	جـ	ـجـ	ـج	
ح	حـ	ـحـ	ـح	
خ	خـ	ـخـ	ـخ	
د			ـد	
ذ			ـذ	
ر			ـر	
ز			ـز	
س	سـ	ـسـ	ـس	
ش	شـ	ـشـ	ـش	
ص	صـ	ـصـ	ـص	
ض	ضـ	ـضـ	ـض	
ط	طـ	ـطـ	ـط	
ظ	ظـ	ـظـ	ـظ	
ع	عـ	ـعـ	ـع	
غ	غـ	ـغـ	ـغ	
ف	فـ	ـفـ	ـف	
ق	قـ	ـقـ	ـق	
ك	كـ	ـكـ	ـك	
ل	لـ	ـلـ	ـل	
م	مـ	ـمـ	ـم	
ن	نـ	ـنـ	ـن	
ه	هـ	ـهـ	ـه	
و			ـو	
ي	يـ	ـيـ	ـي	

تم تصنيف الحروف العربية لغويا إلى صنفين [12]:

1) الحروف الزائدة: سميت هذه الحروف بالزائدة وهـي (ا ' ل، و، ت، س، م، ن، ه، ي)' يجمعها قولهم سألتمونيها' لأنها تفيد في تكوين صيغ قواعديه مختلفة من الاسـم والفعـل بتغييرات تطرأ على الجذر الثلاثي للكلمة.

مثال: "م" هي حرف زائد في كلمة "مشغل".

2) حروف الجذر: الصنف الآخر من الحروف العربية تكون مجموعة حروف الجذر، وهذه الحروف لا تتدخل في الصياغة القواعدية للكلمات كسابقاتها ولكنها تكون فقط جذر الفعل، والمهتمون بقواعد اللغة العربية يأخذون دائما كلمة "فعل" مثالا للجذر الثلاثي للفعل. يمكن الإشارة هنا الى أن الحروف الزائدة قد تكون حروف من جذر الكلمة ولكن حروف الجذر لا يمكن أن تكون حروفا زائدة لان مجموعة الحروف الزائدة هي فقط المذكورة في الفقرة 1.

مثال:

الجذر	صيغة الفعل المزيد	
	أفعل	فاعل
سعد	أسعد	ساعد

أما الأسماء في اللغة العربية فقد ذكر النحاة صنفين من الأسماء وهما:

1) الأسماء الثابتة: هي الأسماء التي لا تشتق من الجذر (فعل) كما أنها لا تنتج كلمات أخرى وهي أسماء العلم.

مثال ذلك: موسى، احمد..

2) الأسماء المشتقة: وهي الأسماء المشتقة من جذر الفعل وهي أسم الفاعل، أسـم المفعـول، أسماء الزمان والمكان، أسماء الأدوات والصفة المشبهة بالفعل.

والأسماء المشتقة تنشأ من جذر الفعل :

* أما بالحركات: جهر ⟵ جهر
* أو بحشر بعض الحروف الزائدة بين حروف جذر الكلمة

مثال ذلك: كَتَب ⟵ كِتاب

هذا الحشر للحروف ليس اختياريا ولكنه يلتزم بصيغ كتابة كل عائلة من هذه الأسـماء التـي حددها النحويون . لذلك فأن "كِتاب" هي على وزن "فِعال".

4. التحليل الإحصائي للنص العربي

العينات المستخدمة في بناء الذخيرة هي مليون كلمة الكترونية خزنت في ملف من نوع Text، حيث تم جمع نصوص عربية مطبوعة في مواضيع مختلفة (رياضية، علمية، اجتماعية، مجلات، أخبار، ثقافة، ... الخ) لضمان ورود أكبر عدد من الكلمات العربية في ذلك الملف واستخدمت كقاعدة معلومات لانجاز الإحصائية.

تم قراءة الملف حرفا حرفاً وألغيت الفراغات البينية والرموز الخاصة التي ترد خلال النصوص، ثم تم حساب ترددات الحروف المتتابعة داخل هذا الملف ونتجت الإحصائية في (الجدول-2) والتي ستستخدم في تصويب الكلمة الخاطئة.

جدول (2): نتائج التحليل الإحصائي للنص العربي

	حرف بدون نقطة	حرف بنقطة في أعلى الحرف	حرف بنقطة في أسفل الحرف	حرف بنقطتين في أعلى الحرف	حرف بنقطتين في أسفل الحرف	حرف بثلاث نقاط
حرف بدون نقطة	49514 (سد)	9819 (دف)	3534 (حج)	5361 (ست)	4808 (سي)	1372 (حث)
حرف بنقطة في أعلى الحرف	6795 (ڤص)	1081 (ڤخ)	577 (نب)	1449 (نق)	3204 (ڤي)	92 (ڤث)
حرف بنقطة في أسفل الحرف	4889 (بد)	673 (بز)	155 (بب)	475 (بت)	901 (بي)	109 (بث)
حرف بنقطتين في أعلى الحرف	5441 (قص)	1003 (قز)	517 (تب)	637 (تق)	565 (قي)	213 (قث)
حرف بنقطتين في أسفل الحرف	5079 (يد)	1267 (ين)	447 (يب)	907 (يت)	949 (يي)	281 (يش)
حرف بثلاث نقاط	1522 (شع)	195 (شف)	82 (شب)	98 (ثق)	423 (شي)	0

5. التحليل المورفولوجي للمفردات

الصرف أو "المورفولوجي" هو فرع اللسانيات الذي يتعامل مع البنية الداخلية للكلمات، من حيث تكوين عناصرها الأولية، والتفاعل بين هذه العناصر، وكلمة "مورفولوجي" – وتعني باليونانية دراسة الشكل – لا تعطي دلالة دقيقة لمهمة الصرف داخل المنظومة اللغوية، وهي المهمة التي لا تقتصر على الأمور المتعلقة بـ "شكل" الكلمات بل تتعداها إلى دراسة الصلة بين مباني الكلمات ومعانيها وخصائصها النحوية والفونولوجية[1]. وقد تعامل هذا البحث مع بنية الكلمة فقط.

لغرض تكامل التحليل في عملية تمييز الكلمات، فأن قاعدة المعرفة المستخدمة في ذلك هي قواميس تتضمن قائمة بكلمات صحيحة توافق المجال الذي ينفذ فيه النظام. يتحقق هذا التكامل إذا تم تنظيم القاموس بطريقة تسهل استخدامه. وتم في هذا النظام بناء ثلاث قواميس[13]:

- قاموس بأسماء الآلات
- قاموس بالأسماء المشتقة
- قاموس بأسماء الأفعال

فكل كلمة لا تكون موجودة في هذه القواميس تعتبر خاطئة، وهنا تبدأ عملية المعالجة لتصويبها.

6. التطبيق العملي

تتوقف عملية تصحيح الأخطاء على نوع الخطأ المحتمل، والذي يشمل: القلب المكاني، وإبدال الحروف، وحذفها، وإضافتها الممكنة، وذلك باستبدال الحروف المنقلبة بتلك التي تشابهها، وتعويض الحروف المحذوفة بتلك المسموح بها داخل سياق الكلمة. يتم اختبار صحة الافتراضات التي يجريها المصحح الآلي بالطريقة نفسها التي يتم بها اكتشاف الأخطاء [1].

تتكون الكلمة التي يتم تصويبها من (n) من الحروف المتتابعة ويطلق على مجموعة الحروف هذه (*n-gram*). حيث إن n تكون 1، 2، 3، أو 4 فإذا كانت $n = 1$ فأن *n-grams* يسمى *uni–grams* أو *mono grams* ; وعندما $n = 2$ يسمى *diagram* أو *bigrams* ; وكذلك إذا $n = 3$ فيسمى *trigrams* الخ.

بصورة عامة، فان التقانات التي تستخدم *n-grams* تختبر كل مقاطع الكلمة المدخلة حيث يتم البحث عنها في الجدول الذي يتضمن إحصائيات تتعلق بـ *n-grams* الأكثر ورودا.

الكلمات التي لا يوجد لها *n-grams* في الجدول الإحصائي فان احتمال كونها خاطئة يكون كبيراً، وإذا لم تظهر في القاموس فذلك يعني أنها خاطئة، وهنا تبدأ المعالجة لتصويبها[12].

تم تحديد المجال الرياضي في النصوص المطروحة للمعالجة في هذه المرحلة لغرض تحجيم مديات المصطلحات التي تخزن في القواميس، هذه القواميس هي:

أسماء الآلات: مثلا (كرة، مضرب، سلة، ...).

الأسماء المشتقة: مثلا (حكم، لاعب، حامي هدف، ...).

الأفعال: مثلا (لعبوا، يلعب، يطرد، يحكم، طرد، يهدف، ...).
حيث تم استخدام برنامج *Notepad* لخزن المعلومات في الملفات.

تُقرأ جملة للاختبار... وبما أن المجال الذي تم اختياره هو المجال الرياضي فسيتم إدخال جملة رياضية للمعالجة، وتتكون المعالجة من مرحلتين:

6-أ- اكتشاف الخطأ

1) تقطع الجملة المقروءة إلى كلمات بواسطة دالة (strtok) وهي دالة جاهزة في لغة C++ (اللغة المستخدمة لبرمجة خوارزميات البحث) هذه الدالة تقطع الجمل إلى كلمات باعتماد الفراغ والرموز الخاصة (;، (،)، ...) بوصفها فواصل بين الكلمات.
2) إلغاء حروف الجر من الجملة استناداً إلى جدول بحروف الجر لغرض استثنائها من المعالجة.
3) إلغاء البادئات واللواحق مباشرة من كل كلمة وهي (أل ـ التعريف، بـ ـ، يـ ـ ـ ـ، و، الألف والنون، الواو والنون، إلى غير ذلك..) لغرض الحصول على أصل الكلمة.
4) تتم موازنة الكلمة الناتجة مع الكلمات في القواميس.
   أ. إذا كانت الكلمة موجودة في احد القواميس فيعني ذلك أنها صحيحة.
   ب. إذا لم يعثر على الكلمة في القواميس فتعد كلمة خاطئة وتحتاج إلى تصويب. تكون الموازنة باتجاهين من اليمين إلى اليسار ومن اليسار إلى اليمين لغرض الحصول على اكبر تطابق مع كلمات القاموس.
5) تقطع الكلمة غير الصحيحة إلى الحروف المكونة لها ويتم موازنتها مع حروف كلمات القواميس لاكتشاف موقع الخطأ، ويعتمد هنا على ASCII Code للتعامل مع الحروف في كل مراحل المعالجة.

6-ب- تصويب الكلمة الخاطئة

بعد حذف البادئات واللواحق من الكلمة الخاطئة والتأكد من أنها ليست حرف جر تدخل الكلمة إحدى المعالجات الآتية:-

1. <u>تصويب الحرف الأول أو الأخير من الكلمة</u>:

بعد أن يتم الحصول على أعلى نسبة تطابق للكلمة تحت الاختبار مع أحدى كلمات القاموس يجب معرفة موقع الحرف المفقود فإذا كان الحرف المفقود هو الحرف الأول فيتم أخذ أول حرف من الكلمة التي حصل معها أعلى تطابق ويلصق مباشرةٍ في بداية الكلمة، وإذا كان الحرف الأخير هو المفقود يلصق الحرف المفقود في نهاية الكلمة.

مثال-1 : اعب ⟵ لاعب

مثال-2 : جزا ⟵ جزاء

2. تصويب حرف وسطي من الكلمة:

بعد أن يتم الحصول على أعلى نسبة تطابق للكلمة تحت الاختبار مع أحدى كلمـات القاموس يجب معرفة موقع الحرف الخاطئ فيلغى ويترك موقعه خاليا ثم الرجوع إلى الخلف خطوة واحدة للتعرف على الحرف قبل الملغي ثم باستخدام نتائج الإحصائية المنجزة في هـذا البحث (الجدول-1) يتم تقدير الحرف البديل ويضاف إلى الكلمة. قد يتحقق أكثر من خيـار صائب مما يستوجب تدخل المستفيد لاختيار الكلمة الأصوب ضمن سياق الجملة.

مثال: كرة الطاوئة ⟵ كرة الطاولة

إذا كان الخطأ في حرفين من الكلمة فيؤخذ احتمال أن تحتاج إلى *قلب مكاني* بان تبدل مواقع الحرفين ثم توازن مع كلمات القاموس. إذا لم يظهر تطابق تعاد الحروف المقلوبة إلى مواقعها ثم يلغى احد الحرفين الخاطئة ثم يعالج الخطأ كما في (2). فإذا لم يظهر تطابق تعاد العمليـة نفسها بعد إلغاء الحرف الآخر.

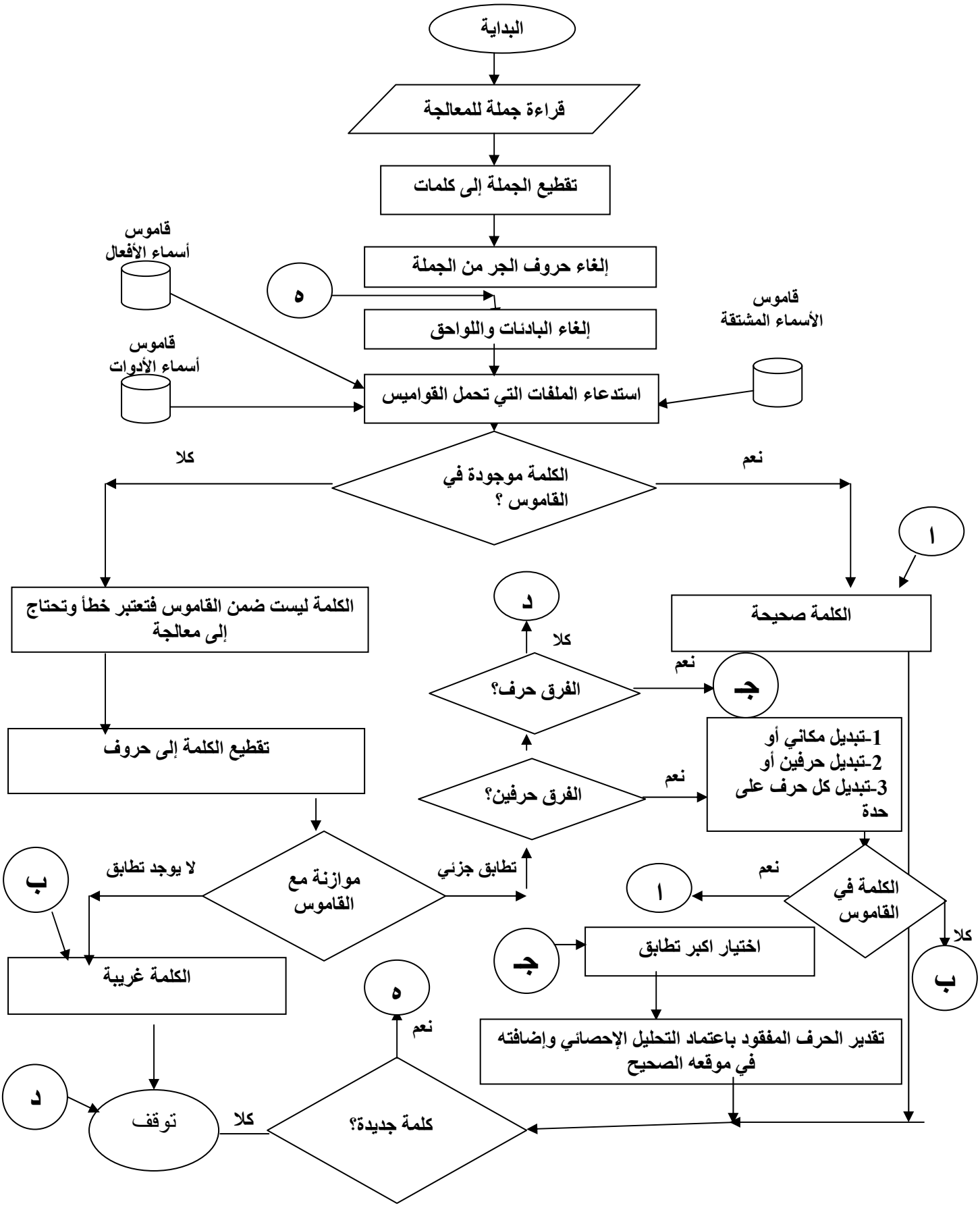
3. قد يكون الحرف الخاطئ بديلا عن حرف يشبهه على لوحة المفـاتيح، فتبـدل 'فـ' بالحرف 'قـ'، و 'نـ' بالحرف 'تـ'، و 'ر' بـالحرف 'ز' أو 'و'،... والعكـس بالعكس، فان هذه الأخطاء شائعة عند إدخال النصوص.

مثال: قفز الزاتة ⟵ قفز الزانة

تتم الموازنة مع كلمات القاموس بعد كل تبديل وإذا لم يحصل تطـابق فـالقرار بـان الكلمة غريبة.

عينات الاختبار عددها مفتوح تحده كلمات القاموس فكلما كبر حجم القـاموس أمكـن زيـادة عينات الاختبار، فقد تم اختبار الخوارزمية المقدمة على (50) عينة في المجـال الرياضـي تباينت فيها الأخطاء وتنوعت طرق التصويب لها فقد تم تصويب (42) كلمة منها وكانت نسبة التصويب 85%.

يوضح المخطط الانسيابي المبين في (الشكل-1) كافة مراحل المعالجة.



(الشكل-1) معالجة التدقيق الإملائي

7. الاستنتاجات وإمكانية التطوير

من المعروف بأنه لا يوجد أي نظام نوع ***OCR*** يحقق نجاحاً في التمييز تـصل إلـى 100%. يعرض هذا البحث طريقة للتدقيق الإملائي للكلمات العربية المطبوعة مـع محاولـة لتصحيح الخطأ وذلك باستخدام تحليل مورفولوجي وتحليل إحصائي للكلمات العربية. التحليـل الإحصائي يستند إلى كم هائل من الكلمات جمعت في ملف واحد بحجم مليون كلمة الكترونية. وثبتت أهمية الإحصائية المنجزة في البحث في إنجاح مرحلة تصحيح الأخطاء التي قد ترد في نصوص بعض الكلمات. الأخطاء التي تم معالجتها في هذا البحث تشمل القلب المكاني، وإبدال الحروف، وحذفها، وإضافتها الممكنة، وذلك باستبدال الحروف المنقلبة بتلك التـي تـشابهها، وتعويض الحروف المحذوفة بتلك المسموح بها داخل سياق الكلمة. تم اختبـار الخوارزميـة المقدمة على (50) عينة في المجال الرياضي تباينت الأخطاء فيها وتنوعت طرق التـصويب لها فكانت نسبة التصويب 85%.

يمكن تطوير هذا البحث بإضـافة تحليـل قواعـد (Syntax Analysis) وتحليـل معنـى (Semantic Analysis).

المراجع

1) د. نبيل علي، (1988)، "اللغة العربية والحاسوب"، دراسة بحثية، منـشورات تركـي العريض.

2) Jurafsky, D. and Martin, J. (2000), "Speech and Language Processing", Prentice Hall.

3) Tseng, Y., and Oard, D. (2001), "Document Image Retrieval Techniques for Chinese", In Symposium On Document Image Understanding Technology, pp. 151-158, Columbia, MD.

4) Darwish K., Hassan H., and Emam O., (2005), "Examining The Effect Of Improved Context Sensitive Morphology On Arabic Information Retrieval", Proc. Of ACL Workshop on Computational approaches to Semitic-Languages", Ann-Arbor, 2005.

5) Ahmad H., Sara N., and Hany H., (2008), "Language Independent Text Correction Using Finite state Automata", Proc. Of Int. Joint Conference on Natural Language Processing (IJCNLP08).

6) A. Amin and J.F. Mari, (1989), "Machine Recognition and correction of Printed Arabic Text", IEEE trans. On Systems, Man and Cybernities V 19, No.5, pp 1300-1306.

7) Sari T., and Sellami M., (2002), "MOrpho-LEXical Analysis for Correcting OCR-Generated Arabic Words (MOLEX)", Proc. Of 8th International Workshop on Frontiers in Handwriting Recognition, PP.461-466.

8) Walid M. and Kareem D., (2008), "Effect of OCR Error Correction On Arabic Retrieval", Information Retrieval (11), PP: 405–425.

9) Al-Talib Ghayda, (2006), "Fuzzy Logic Based Arabic Optical Character Recognition", Ph.D. Thesis, College of Computers and Mathematics Science, University of Mosul.

10) Miled H., Oliver C., et. al., (1997), "Coupling Observation / Letter For Markovian Modelisation Applied To The Recognition Of Arabic Handwriting", IEEE, pp: 580-583.

11) Khedher M. Z. and Abandah G. A., (2002), "Arabic character recognition using approximate stroke Sequence", 3rd Int. Conf. on Language Resources and Evaluation (LREC), Workshop.

12) Toufik S. and Mokhtar S., (2005), "Correcting Arabic OCR Output by Morphological Analysis of Words", International Conference on Machine Intelligence (ICMI), Tunisia.

13) Wells C., Evett L., et. al., (1990), "Fast Dictionary Look-Up for Textual Word Recognition", Pattern Recognition, Vol. 23, No. 5, PP 501-508.