

نظام حاسوبي لتشكيل النص العربي
Automatic Arabic Text Diacritizer

التقرير الفني النهائي
Final Report

إعداد

د. منصور بن محمد الغامدي*

د. محمد خورشيد*؛ د. مصطفى الشافعي**؛ د. فايز الحرقان*؛

د. أبو أوس الشمسان***؛ د. محمد الكنهل*؛

سعد القحطاني****؛ سيد زيشان مظفر**؛ ياسر التويم*؛ عدنان يوسف**؛ حسني المحتسب**

* مدينة الملك عبد العزيز للعلوم والتقنية

** جامعة الملك فهد للبترول والمعادن

*** جامعة الملك سعود

**** وزارة الدفاع والطيران

*Mansour Alghamdi**

Muhammad Khursheed, Mustafa Elshafei**, Fayz Alhargan*,*

Muhammed Alkanhal, Abu Aus Alshamsan***, Saad*

*Alqahtani****, Syed Zeeshan Muzaffar**, Yasser Altowim*, Adnan*

*Yusuf**, Husni Almuhtasib ***

**KACST*

***KFUPM*

****KSU*

*****MODA*

١٤٢٧/٥/٢٢ هـ

الموافق

٢٠٠٦/٦/١٨ م

الملخص:

يهدف هذا المشروع إلى دراسة نظام التشكيل في اللغة العربية ومن ثم كتابة برمجيات تقوم بتشكيل النص العربي آلياً بهدف دمج هذه البرمجيات في نظم حاسوبية أخرى كالناطق الآلي والتعرف الآلي على الكلام العربي ومحركات البحث وغيرها. وساهم في المشروع عدد من الباحثين والمطورين من مدينة الملك عبد العزيز للعلوم والتقنية وجامعة الملك فهد للبترول والمعادن وجامعة الملك سعود ووزارة الدفاع والطيران. وقام الفريق طيلة فترة المشروع بتنفيذ عدد من البحوث والمحاولات لبناء نظم حاسوبية تتولى عملية تشكيل النص العربي، فتمكنوا من بناء ثلاثة نظم حاسوبية مختلفة تقوم بتشكيل الآلي للنص العربي: النظام الأول، نظام التشكيل بأدوات ماركوف الخفية (سيطلق عليه "النظام الأول" في التقرير)، والنظام الثاني: نظام التشكيل بالفيتربي (وسيطلق عليه "النظام الثاني")، والنظام الثالث، نظام التشكيل المستقل (وسيطلق عليه "النظام الثالث"). ويعتبر النظام الثالث إنجازاً يسجل للمدينة من حيث فكرته وأدائه، إذ وصلت نسبة الصحة في تشكيله ٨٧% لأي نص عربي ولجميع حروف الكلمة. وقد تم إيداعه في الإدارة العامة لبراءة الاختراع بالمدينة في ١٥/٥/٢٧هـ. كما نتج عن المشروع ورقتان علميتان منشورتان.

الكلمات المفتاحية:

النص العربي، علامات التشكيل، الآلي، الحروف، اللغة العربية

Abstract:

The objective of this project is to study the diacritization system in Arabic text and hence build a system that would be able to diacritize Arabic text automatically. Such a system can be integrated into other systems such as text-to-speech and speech-to-text systems. Since this area of research was new at KACST, researchers from different academic and research institutions were invited to participate in the project. The team included researchers from King Abdulaziz City for Science and Technology, King Fahad University for Petroleum and Minerals, Ministry of Defense and Aviation and King Saud University. The team investigated different approaches for diacritizing Arabic text and they built three systems: Automatic Diacritizer of Arabic Text Using Hidden Markov Model (System I), Automatic Diacritizer of Arabic Text Using Viterbi (System II) and An Independent Diacritizer of Arabic Text (System III). System III is considered an achievement for the research team for several reasons. One of them, it is independent. The second, its performance is higher and faster than the other systems. The system was submitted to the General Directorate of Patents at KACST for Patent on 11th of June 2006.

Keywords:

Arabic text, diacritic, diacritization, vowel, grapheme, automatic

المحتويات

٣	المقدمة:	١
٧	منهج البحث:	٢
٩	النتائج:	٣
٢٨	الخلاصة:	٤
٣٠	الملاحق:	٥

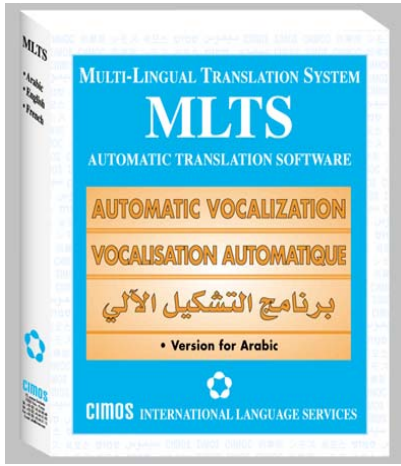
١ : المقدمة:

تظهر الكتابة العربية المعاصرة على شكل حروف دون علامات التشكيل، هذا ما يقرأه العربي في الكتب والصحف والإعلانات وعلى شبكات الإنترنت. وهذا يعني أن الخط العربي المعاصر يفتقد للرموز التي تمثل الصوائت. وقد لا يبرز ذلك إلا عند المقارنة بكلمات من لغات أخرى. فلو استبعدنا الصوائت من كلمات إنجليزية مثل: read. ride. rode. red. rod. rude، لأصبحت جميع هذه الكلمات كلمة واحدة وهي rd ولأصبح من المستحيل معرفة أي منها تعني لو كتبت بمفردها. وفي حقيقة الأمر أن هذا ما يحدث عند قراءة معظم الكلمات العربية. إلا أن القارئ العربي يقوم بإكمال النقص في الصوائت وذلك بإضافة الحركات المناسبة للحروف حسب معرفته باللغة العربية وحسب سياق الجملة. هذه العملية تتم تلقائياً ودون الإحساس بها. فكل متحدث عربي يقوم بها دون الشعور بالتفكير فيها. ومما يميز نظام كتابة اللغات السامية بما فيها العربية الاختزال ومن ذلك حذف الرموز التي تمثل الصوائت .

إلا أن عدم وجود علامات التشكيل على الحروف العربية يسبب معضلة ليست باليسيرة لكثير من البرمجيات المعاصرة ومنها الناطق الآلي. فلا يمكن لأي نظام نطق آلي أن يقوم بعملية النطق دون تشكيل للحروف. فعند ورود أية كلمة عربية لا يمكن لنظام الناطق الآلي معرفة علامات التشكيل التي تستوجب وضعها على الحروف المكونة لها. لهذا لا بد من توفير مشكل آلي ليقوم الناطق الآلي بمهمته. وتمثل علامات التشكيل فيما تمثله الصوائت التي لها دور أساسي في الخصائص الفيزيائية لموجات الكلام المنطوق. ولهذا تصبح عملية النطق الآلي عملية مستحيلة دون التشكيل. كما أن للتشكيل أهمية في عمليات حاسوبية أخرى كما في محركات البحث والنقل الكتابي للأسماء من وإلى اللغة العربية والتعرف الآلي على الكلام العربي وغيرها.

ولهذا ظهرت محاولات عديدة لتشكيل الحرف العربي من قبل شركات مختلفة بهدف استخدام المشكل الآلي في نظم حاسوبية أخرى بما في ذلك الناطق الآلي. فقدمت الشركة الهندسية لتطوير نظم

الحاسبات¹ RDI المشكّل الآلي للنص العربي RDI ArabDiac© 2.0 وتقوم التقنية المبني عليها هذا المنتج على المحلل الصرفي: ٣,٠٠ . RDI ArabMorpho© وتبلغ سرعة التشكيل الآلي للنصوص العربية ١٠٠ كلمة/ثانية على جهاز PIII 750 MHz يحتوي على ذاكرة ٦٤ ميغا بايت ويعمل على نظام تشغيل ميكروسوفت ويندوز. كما أن دقة التشكيل الآلي تزيد عن ٩٥% مقيسةً على مستوى الكلمات. هذه التقنية منفذةً بالكامل باستخدام ANSI C++ مما يمكن من تجهيزها للعمل تحت أي نظام تشغيل (ويندوز، يونيكس، ماكينتوش، ...) بسهولة ويسر. نورد هذه المعلومات حسب ما تذكره الشركة المصنعة إلا أنها لم تختبر من جهة مستقلة. وقامت شركة سيمو² الفرنسية بإنتاج نظام للتشكيل الآلي للنص العربي، وقد تم شراء هذا البرنامج وتجريبه وكانت نسبة التشكيل الصحيح تساوي ٧٠% تقريباً على مستوى الكلمة وعليه حماية إذ لا يسمح باستخدامه في أكثر من جهاز واحد.



يضيف هذا البرنامج علامات التشكيل إلى النص العربي، وهو على ثلاث نسخ: الأولى تعمل مستقلة على جهاز الحاسب الآلي، الثانية تندمج في نظم أخرى عبر API، الثالثة تتركب على الخادم لتعمل عبر الشبكة العالمية.

Diacritizer



كما قامت "الشركة العالمية"³ بتطوير نظام تشكيل يبدو أنه الأسرع والأدق (حسب قول الشركة المنتجة) حيث يقوم بالتشكيل بسرعة عالية ودقة تصل إلى ٩٨% ويعطي الخيار بين تشكيل أو آخر الكلمات (الإعراب) أو عدمه. هذا النظام هو جزء من برنامج صخر "أدوات المكتب" office tools.

¹ <http://www.rdi-eg.com>

² <http://www.cimos.com>

³ <http://www.aramedia.com/diacritizer.htm>

<http://www.sakhr.com>

ولقد قامت شركات أخرى بجهود في هذا المضمار لخدمة برمجيات لها علاقة بالتدقيق الإملائي والنطق الآلي ومعالجة النصوص العربية. ومن هذه الشركات: شركة "أي بي إم" ، و "إنفو آراب" ، و "كولتك".⁴

ويؤخذ على منتجات الشركات المذكورة آنفا فيما يتعلق بالمشكل الآلي التالي:

١. قليل من الشركات المذكورة توفر برنامج مشكلها الآلي مستقلا بذاته. والشركتان الوحيدتان التي يمكن شراء المشكل الآلي منهما مستقلا هما شركة "سيمو" والشركة الهندسية لتطوير نظم الحاسبات. وحتى بعد شراء المشكل الآلي فإنه محمي لدرجة أنه من الصعب دمجه في نظم أخرى أو تحميله على حاسب آخر ناهيك عن تطويره أو تعديل برمجياته.

٢. المشكل الآلي المتاح مغلق المصدر، ومن سلبيات البرمجيات من هذا النوع أنها لا تفصح عن الكثير من الخصائص والأسس التي بنيت عليها مما يجعل مجال الاستفادة منها أو تطوير فكرتها غير ممكنة، إضافة إلى ذلك فإن هذا النوع من البرمجيات لا يتيح للمتخصصين والمبرمجين إمكانية تطوير شفرته code أو تعديلها ليتلاءم مع نظم وبرمجيات وتطبيقات حاسوبية أخرى .

وللأسباب المذكورة أعلاه حول البرمجيات التي تنتجها الشركات، فقد قام عدد من الباحثين بجهود فردية بهدف تطوير برمجيات للتشكيل الآلي. ومن هذه الجهود ما يتعلق باللغة العربية ومنها ما يتعلق بلغات أخرى ولأهداف لغوية مختلفة ولكنها ذات علاقة بحوسبة اللغة والتعامل مع النص اللغوي. مما يجعل من الممكن الاستفادة من هذه التقنيات والأساليب في تشكيل النص العربي. فقد قام Chelba and Jelinek (2000) بعرض طريقة للتحليل الآلي لمكونات النص من كلمات في اللغة الإنجليزية مستخدما احتمالية تسلسل الكلمات والارتباطات الشجرية بينها. لم يعرض الباحثان نتائج بحثهما، إلا أن أصل البرنامج source code ضمن الورقة⁵. ويمكن الاستفادة من نتائج هذه الورقة في تشكيل الكلمات العربية ولو أن الاستفادة محدودة وذلك لكون الكلمات العربية تتمتع بحرية التسلسل وهذا عكس مفردات اللغة الإنجليزية التي تخضع لنظام تسلسل ثابت يقوم على: الفاعل + الفعل + المفعول به.

وقد قام Goweder وآخرون (٢٠٠١) بعمل إحصاء لتكرار الكلمات الواردة في جريدة الحياة. وكان هناك معالجة مسبقة للنصوص قبل التحليل الإحصائي وذلك لتوحيد الكلمات التي تكتب بالطريقة نفسها أو استبعاد الكلمات الغير صحيحة. وقد أبرز البحث النتائج الإحصائية المبينة في الجدول التالي:

هذا يعني أنه كلما قل عدد الكلمات في النص كلما كانت احتمالية تكرار الكلمات منخفضة وأنه كلما زاد عدد المفردات كلما ارتفعت النسبة. ورغم أن هذه الظاهرة طبيعية في النص اللغوي إلا أن البارز هنا هو أن اللغة العربية تتمتع بنسبة منخفضة لتكرار الكلمات في النص، وتختلف النتيجة هنا عن

⁴ www.ibm.com

⁵ www.isisintl.com/Linguistics

⁶ www.coltec.net

⁷ www.idealibrary.com

الإنجليزية التي تكون فيها نسبة التكرار أعلى حيث تصل - على سبيل المثال- إلى خمسة عندما يكون عدد مفردات النص ١٠.٠٠٠ كلمة^٨

عدد الكلمات	١٠٠	١٠.٠٠٠	١٠٠٠.٠٠٠	١٠.٠٠٠.٠٠٠	٢٠.٠٠٠.٠٠٠
نسبة التكرار	١	٢	٨	٢٥	٤٢

وقام Gal (2002) باستخدام HMM في تحليل النصوص العربية المشكلة ثم بناء أنموذج للتشكيل الآلي. والنص العربي المستخدم في هذه البرمجيات هو نص القران الكريم. وقد بلغ عدد الكلمات المستخدمة ٩٠.٠٠٠ كلمة. استخدم ٩٠٪ من النص للتدريب و ١٠٪ المتبقية لاختبار النظام. ومن نتائج هذا البحث: (١) أن ٧٠٪ من الكلمات لها أكثر من معنى واحد إذا لم تشكل؛ أي لها أكثر من تشكيل. (٢) استخدم الباحث نظام الكلمتين المتتاليتين ونظام الكلمة الواحدة. (٣) نتيجة اختبار النظام: نسبة الكلمات صحيحة التشكيل ٨٦٪، وأن ٨٪ من نسبة الخطأ تعود إلى كون الكلمات لم تكن موجودة في نص التدريب. (٥) توصية باستخدام محلل صرفي يكون داعما للتشكيل. ويمكن استخدام المحلل الصرفي الذي أنجزته شركة زيروأكس الذي يذكر المؤلف بأنه متوفر مجانا. (٦) توصية باستخدام الحروف إضافة إلى الكلمات. (٧) توصية باستخدام نظام الثلاث كلمات. (٨) توصية بوضع معجم للكلمات ثابتة التشكيل. (٩) توصية بزيادة النصوص المشكلة المستخدمة في التدريب.^٩

استخدم Kontorovich, Leonid and Daniel (2001) نصوصا مشكلة من العبرية -٣٠.٠٠٠ كلمة- واختبرت بنصوص أخرى -٣٠٠٠ كلمة- كلا النصين من التوراة. واستخدم الباحثان ثلاث طرق للتشكيل: (١) قاموس مستقل عن السياق: وتقوم هذه الطريقة بتشكيل كل كلمة منفردة بناء على أعلى تكرار حصلت عليه في التدريب على التشكيل. مثال: لو وردت كلمة "سمع" بالتشكيل "سَمَع" ٣٠ مرة، و بالتشكيل "سَمَع" سبع مرات، فإن النظام سيشكلها دائما "سَمَع" في أي مكان يجدها. (٢) استعمال وظيفة الكلمة: صنفت الكلمات حسب وظيفتها ويتم تشكيلها بناء على الوظيفة، فلو كانت الكلمة "علم" مصنفا على أنها "فعل ماض" فإن التشكيل يكون "عَلِم" ولو كانت مصنفة على أنها "فعل مبني للمجهول" فإنها تشكل "عَلِم"، وذلك بناء على المعطيات المدونة عن كلا النصين؛ نص الاختبار ونص التدريب. (٣) استخدام أنموذج هيدين ماركوف: يعتمد هذا الأنموذج على وضع الكلمة في الجملة وذلك بتكوين جدول لكل كلمة وما يليها ويسبقها من كلمات عند التدريب، وعند التشكيل

⁸ www.elsnet.org/arabic2001/goweder.pdf

⁹ www.cs.um.edu.mt/~mros/WSL/papers/gal.pdf

الآلي يتولى النظام البحث عن الكلمات السابقة واللاحقة بغض النظر عن قربها المباشر من الكلمة المراد تشكيلها. وكانت النتائج متمثلة في الجدول التالي¹⁰:

الطريقة	نتائج التدريب	نتائج الاختبار
١	%٧٤	%٧٧
٢	%٧٦	%٧٩
٣	%٨٣	%٨١

استخدم (1998) Mustafa أربع خوارزميات للبحث عن الكلمات العربية سواء كانت مشكلة أم غير مشكلة. ووجد الباحث أن خوارزميات Boyer-Moore-Horspool algorithm هي أفضل برمجيات استخدمها من حيث سرعة البحث والتصحيح الإملائي. وقد استخدم الباحث في مشروعه نصوصاً مشكلة وشبه مشكلة (٦% من الحروف مشكلة) وغير مشكلة. ووجد أن طول النص والتشكيل عاملين مهمين في إبطاء سرعة البرمجيات.

يمكن الاستفادة من نتائج هذه الورقة في عملية معالجة النص فيما قبل التشكيل وذلك لمعالجة الأخطاء والاختلافات الإملائية والتشكيل الجزئي الذي قد يصاحب النص المراد تشكيله.

وقد استعرض (2002) Smrž and Zemánek استخدام الأسلوب الشجري في وضع قوانين لتحليل النص العربي، والذي يرا الباحثان أنه قد يساعد في تشكيل الكلمات العربية¹¹.

ونشرت في الآونة الأخيرة أوراق علمية تتعلق بالتشكيل الآلي للنص العربي (Ananthakrishnan, et al 2005; Dimitra and Kirchho 2004; Nelken and Shieber 2005). إلا أنها لم تخرج عن إطار استخدام أنموذج ماركوف الخفي مع بعض الإضافات كإدخال الإشارة الصوتية للكلام لرفع نسبة التعرف على الكلام المنطوق آلياً (Ananthakrishnan, et al 2005; Dimitra and Kirchho 2004) أو إضافة مفردات من المعجم اللغوي للعربية (Nelken and Shieber 2005)، وظلت نسبة التشكيل الصحيح متدنية، حيث كانت جميعها أقل من ٨٩% وتتحدر إلى بداية السبعينات عندما يستخدم أنموذج ماركوف الخفي منفرداً.

٢: منهج البحث:

ومما سبق ذكره في المسح الأدبي فقد بدى للفريق العلمي في بداية الأمر أن الأسلوب الإحصائي باستخدام أنموذج ماركوف الخفي هو أنسب طريقة للتشكيل الآلي للنص العربي. لذا عمدنا منذ بداية المشروع في البحث عن آلية استخدام هذا الأنموذج الذي بني في الأصل للتعرف الآلي على الكلام. فكانت الخطوة الأولى بناء متن لغوي مشكل وذلك بجمع نصوص عربية غير مشكلة ثم الطلب من

¹⁰ www.cs.rhul.ac.uk/colt/nipstext.html

¹¹ <http://ufal.mff.cuni.cz/publications/year2002/pbml-sherds.pdf>

لغويين القيام بعملية التشكيل اليدوي لها وذلك لتكوين ذخيرة لغوية كافية لتدريب محرك التشكيل (أدوات ماركوف الخفية التي تقوم على أنموذج ماركوف الخفي).

وبدأ المشروع بتكوين مجموعتي بحث أساسيتين: المجموعة الأولى، لتطوير محرك تشكيل قائم على أنموذج ماركوف الخفي ويقوم بتطوير هذا المحرك مبرمجين متخصصين ولهم دراية بعمل أنموذج من هذا النوع. المجموعة الثانية، لبناء ذخيرة لغوية مشكلة يساهم في جمعها وتشكيلها نخبة من اللغويين المتخصصين في تراكيب الجمل العربية وما تحويه من صيغ نحوية وصرفية.

كما وضع الفريق العلمي في خطته القيام بمحاولة بناء نظام للتشكيل يكون مستقلا عن أدوات ماركوف الخفية إذا أمكن تحقيق نسبة معقولة من التشكيل الصحيح.

ولتنفيذ العنصر الأول (محرك التشكيل الآلي) قام الفريق في بداية الأمر باستقطاب المتخصصين في هذا المجال ومن سبق لهم التعامل مع أدوات ماركوف الخفية من قطاعات مختلفة. وكان هناك اجتماعات اسبوعية لتحديد آلية توظيف أدوات ماركوف الخفية التي وضعت في الأصل لبناء نظم التعرف الآلي على الكلام. وكان على الفريق اكتشاف مكونات أدوات ماركوف الخفية التي يمكن توظيفها في التشكيل الآلي. وبعد عدة لقاءات ومحاضرات ودورات تعريفية بأنموذج ماركوف الخفي وأدواته توصل الفريق إلى المكونات التي يحتاج إليها من أدوات ماركوف الخفية وكيفية التعامل مع النص العربي.

ووصل الفريق إلى أنه يمكن توظيف أدوات ماركوف الخفية في التشكيل الآلي وذلك بتحويل الكتابة العربية من الحرف العربي إلى رموز بحيث يرمز لعلامات التشكيل بحروف رومانية تسمى "الراصد" observation (الملحق: ١) ويرمز لحروف الكتابة بالأرقام وتسمى "واسمة" lable (الملحق: ٢). ويعطى لكل حرف عربي رقما مختلفا بناء على موقعه في الكلمة (بداية الكلمة ووسط الكلمة ونهاية الكلمة) ويعطى للفراغ بين الكلمات رقما أيضا. وبهذا تكون مجموع الأرقام للنص العربي الغير مشكل هو (١١٠) تعطي صوة مطابقة تماما للكتابة العربية ولكن بالأرقام. هذا التفصيل في مواقع الحروف يساهم في دقة إختيار التشكيل المناسب لكل حرف حسب موقعه من الكلمة. ومن ثم فإن العبارة التالي "مَنْ أَنْتَ" يتم تحويلها ببرمجيات أخرى أعدت لهذا الغرض إلى قائمتين بالصيغة التالية حتى تتمكن أدوات ماركوف الخفية من التعامل معها:

<u>قائمة الواصلات</u>	<u>قائمة الراصدات</u>
non	110
mfa	91
mso	95
non	109
mfa	16
mso	95
mfa	27
non	110

وتسلسل الراصدات أعلاه يمثل الآتي: 110 هي بداية العبارة، و 91 حرف "م" في بداية الكلمة، 95 حرف "ن" في نهاية الكلمة، و 16 حرف "أ" في بداية الكلمة وهكذا حتى نهاية القائمة كما هو مبين في الملحق: ٢. بينما تسلسل الواسمات يمثل: non بدون تشكيل، و mfa المفتحة، و msو السكون، وهاكذا مع بقية التسلسل كما هو في الملحق: ١.

وبهذا يحول النص العربي المشكل إلى سلسلة من الرموز كالمذكورة أعلاه ويتم إدخالها على أدوات ماركوف الخفية للتدريب عليها. بعد أن يتم التدريب يصبح النظام جاهزا للتعامل مع نصوص غير مشكلة وذلك باختيار علامات التشكيل المناسبة لحروفها بناء على التدريب الذي تم له مع نصوص مشكلة. وأعدت مكونات أدوات ماركوف الخفية ذات العلاقة بنظام التشكيل وهيأت للتعامل مع التسلسل الذي يمثل النص العربي بطرق مختلفة رغبة في الحصول على أعلى نسبة ممكنة للتشكيل الصحيح. بالتزامن مع تنفيذ العنصر الأول بدأ تنفيذ العنصر الثاني (الذخيرة اللغوية) فقد تم تشكيل فريق من اللغويين قاموا بجمع نصوص لغوية مكتوبة. ووضعت الضوابط التالية لاختيار النصوص:

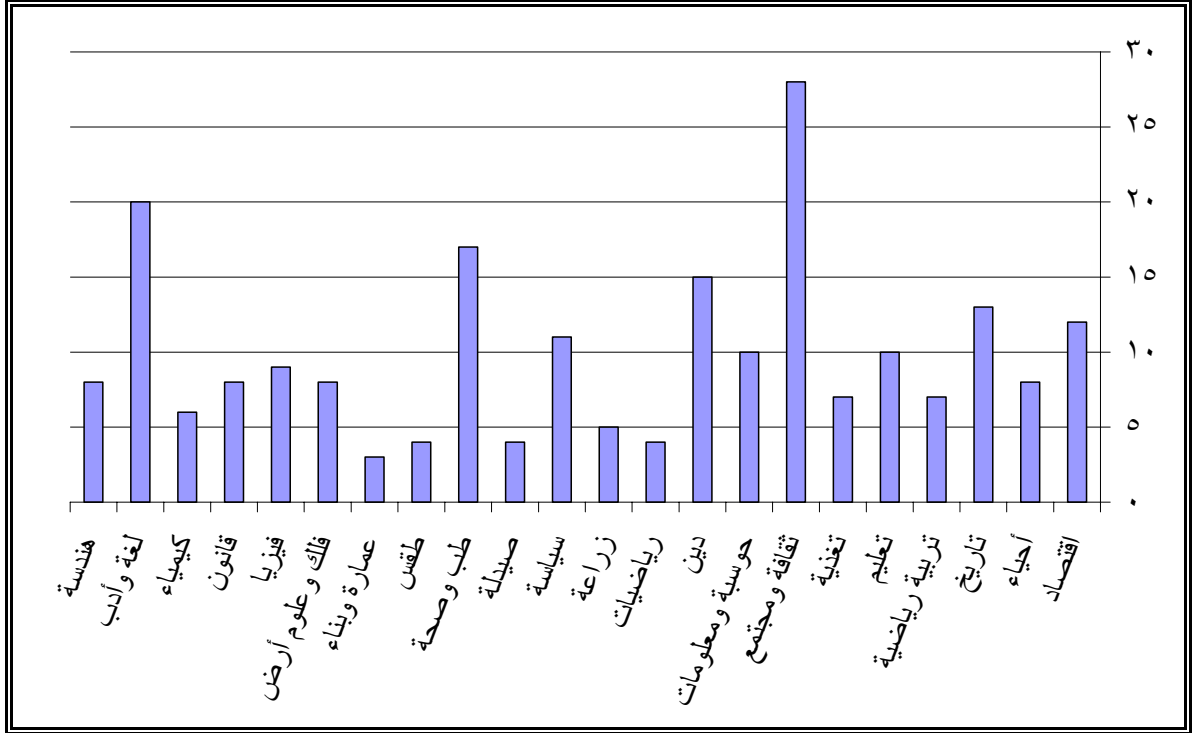
١. أن تكون النصوص منشورة في دوريات متخصصة أو كتب.
 ٢. ألا يؤخذ أكثر من نص للمؤلف نفسه.
 ٣. أن يقع عدد مفردات النص المقتبس بين الألف كلمة والألف والمائة.
 ٤. أن يكون النص خال من الرسومات والرموز والأرقام والأحرف الأجنبية قدر الإمكان.
 ٥. التوازن في اختيار النصوص من مجالات معرفة متنوعة.
 ٦. أن تكون النصوص مكتوبة باللغة العربية الفصحى.
- ثم أدخلت النصوص في الحاسب الآلي وصححت الأخطاء اللغوية متى ما وجدت واستبعدت الصور والأرقام والرموز والحروف الأجنبية إن وجدت في النص وقام الفريق اللغوي بعد ذلك بتشكيل كامل حروفه.

٣: النتائج:

- ٣.١. سنستعرض في البداية الذخيرة اللغوية التي استخدمت في تدريب واختبار نظم المشكل الآلي. فقد استخدم في هذا المشروع ثلاثة ذخائر لغوية مختلفة وهي:
 - ٣.١.١. ذخيرة المشروع:
- وهي الذخيرة التي نفذها فريق المشروع والمشار إليها سابقا. تحتوي هذه الذخيرة على نصوص متوفرة في ٢٣١ ملف نصي، كل منها يحتوي على أكثر من ١٠٠٠ كلمة عربية مشكلة. عند اعدادنا لهذه الذخيرة حاولنا قدر المستطاع تضمين أكبر قدر من المواضيع والمفردات العربية فيها، ولهذا بنظرة سريعة لعدد النصوص ومواضيعها في الشكل: ١. يتضح أن نصوص هذه الملفات تصنف تحت عدة مواضيع كالرياضة والحوسبة والسياسة والدين والاقتصاد والأدب وغيرها من المواضيع. وهذه بعض الإحصائيات عن هذه الذخيرة:

• عدد الأحرف العربية فيها = ١٠٩٥٤٩٦

- معدل عدد الكلمات في الجملة الواحدة = ١٠.٤
- عدد الرموز (أحرف عربية، فراغات، علامات تشكيل) = ٢١٧١٨٦٢
- عدد الجمل الخاطئة في الذخيرة = ٢٨٩
- عدد علامات التشكيل = ٨٤٨٠٩٩
- عدد الجمل الصحيحة = ٢٤٢٧٤
- عدد الكلمات الصحيحة (باستبعاد الجمل الخاطئة) = ٢٥٢٥٤١



الشكل ١. عدد الكلمات (بالآلاف) في كل من المجالات المعرفية التي جمعت تحتها النصوص.

٣.١.٢. القرآن الكريم:

بغرض زيادة عدد المفردات المستخدمة في تدريب واختبار النظام، قمنا بإدراج نصوص القرآن الكريم ضمن البيانات المستخدمة في هذا المشروع. وهذه بعض الإحصائيات عن هذه الذخيرة:

- عدد الأحرف العربية في هذه الذخيرة = ٢٣٨٩٤٦
- معدل عدد الكلمات في الجملة الواحدة = ١١.٦
- عدد الرموز (أحرف عربية، فراغات، علامات تشكيل) = ٤٧٥٩٢٦
- عدد الجمل الخاطئة في الذخيرة من حيث التشكيل = ١٠٤٠
- عدد علامات التشكيل = ١٨٠٧٧٧
- عدد الجمل الصحيحة = ٥٣٠٨
- عدد الكلمات الصحيحة (باستبعاد الجمل الخاطئة) = ٦١٥١١

٣.١.٣. الأسماء العربية:

تزامن هذا المشروع مع مشروع آخر ينفذه المعهد حول الأسماء العربية ورومنتها وحيث قام فريق المشروع الأخير بتكوين قاعدة بيانات للأسماء العربية مشكلة الحروف، فقد استخدمت هذه القاعدة المكونة من ٥٤ ألف اسم في تدريب النظام لنتمكن النظام من التعرف على أكبر قدر من الأسماء ومن ثم التمكن من تشكيلها تشكيلا صحيحا عند وروده في نص غير مشكل. وهذه بعض الإحصائيات عن هذه الذخيرة:

- عدد الأحرف العربية في هذه الذخيرة = ٣٩٩٨٠٩
- معدل عدد الكلمات في الجملة الواحدة = ٢.٦
- عدد الرموز (أحرف عربية، فراغات، علامات تشكيل) = ٧٥٢٣٠٤
- عدد الجمل الخاطئة في الذخيرة = ٠
- عدد علامات التشكيل = ٢٦٥٤٠٠
- عدد الجمل الصحيحة = ٥٤٤٦٣
- عدد الكلمات الصحيحة = ١٤١٥٥٨

ويلاحظ في الإحصائيات السابقة أن هناك بعض الجمل الخاطئة في كل ذخيرة مستخدمة في هذا المشروع. والمقصود بتلك الجمل التي بها خطأ أو أكثر في تشكيل حروف وكلمات هذه الجملة. اكتشفت هذه الأخطاء من قبل برنامج حاسوبي مبسط صمم خصيصاً لهذا الغرض. وبالإضافة إلى اكتشاف الأخطاء فإن من مهام هذه البرنامج أيضاً استبعاد هذه الجمل الخاطئة من الذخيرة لضمان صحة البيانات المستخدمة في تدريب النظام. فإنه بطبيعة الحال لا يستطيع أي برنامج اكتشاف كل الأخطاء الممكنة لكننا بحثنا في عدد من الكتب والمراجع النحوية والصرفية عن بعض الأخطاء الممكن اكتشافها برمجياً، ومن ضمن هذه الأخطاء ما يلي:

- علامة تشكيل تأتي على فراغ.
- علامتا تشكيل تظهران على الحرف نفسه. (مع الأخذ في الاعتبار أن الشدة والحركة التي تليها تعاملان كوحدة واحدة في النظامين الأول والثاني في هذا المشروع).
- شدة تظهر وحيدة على الحرف.
- تتوین لا يظهر في نهاية الكلمة أو قبل ألف في آخر الكلمة.
- بعض التعارض بين حروف معينة وعلامات تشكيل (مشروحة لاحقاً بالتفصيل في الملحق: ٢).

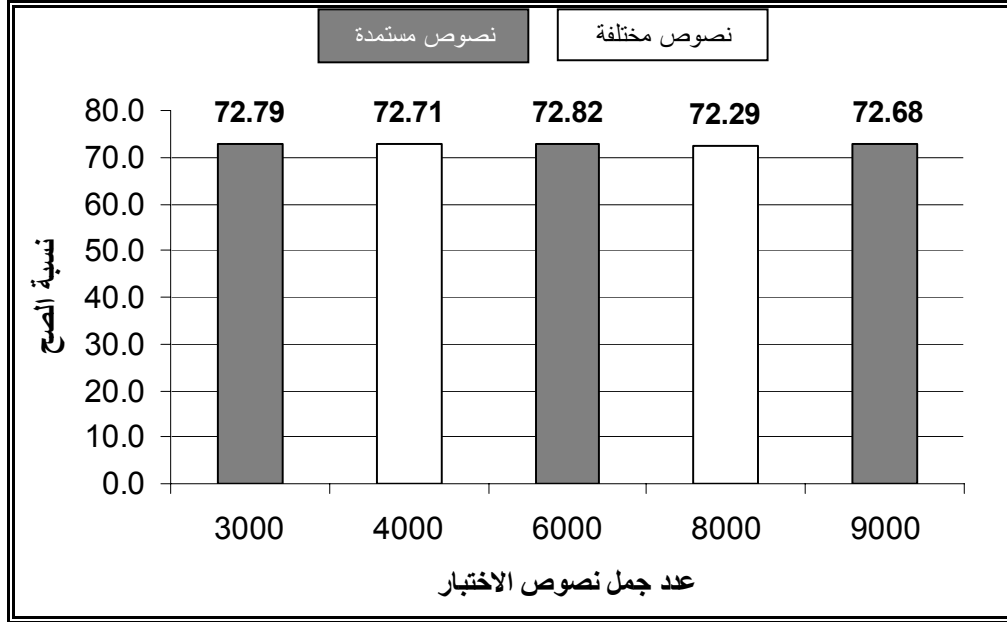
٣.٢. نظم التشكيل

٣.٢.١. نظام التشكيل باستخدام أنموذج ماركوف الخفي (النظام الأول)

استخدام أنموذج التسلسل المفرد للحروف uni-model:

وبعد الإنتهاء من تجهيز الذخائر اللغوية وتهيئة أدوات ماركوف الخفية للتعامل مع التسلسل المقابل للنص العربي كانت أول التجارب المستخدمة في هذا النظام بتدريبه بعشرة آلاف جملة عربية من ذخيرة المشروع ومن ثم استخدام عدة نصوص لاختبار أداء هذا النظام. وقد اقتصر النظام في مرحلته الأولى على ما يعرف بالتسلسل الفردي للحرف uni-gram. وهذه الطريقة تتعامل مع كل حرف كوحدة مستقلة عن الحروف المجاورة الأخرى سواء في التدريب أم التشكيل.

أما النصوص المستخدمة في الاختبار فكانت إما مستخرجة (مستمدة) من نصوص التدريب أو مختلفة عنها تماما. أما نسبة الصحة من هذه التجارب فهي موضحة في الشكل: ٢. كما يوضح الشكل: ٢ كذلك أن نسبة الصحة متقاربة بشكل كبير بغض النظر عما إذا كانت نصوص الاختبار من ضمن نصوص التدريب أم لا. كما تبين هذه البيانات والبيانات في جدول: ١ أن ذخيرة لغوية مكونة من عشرة آلاف جملة تكفي لتدريب النظام على التشكيل الآلي.



الشكل ٢: . نسبة التشكيل الصحيح عند استخدام عشرة آلاف جملة من ذخيرة المشروع في تدريب نظام uni-model واختباره بنصوص أيضا من ذخيرة المشروع تكون إما مستمدة من جمل التدريب (نصوص مستمدة) أو غير مستمدة (نصوص مختلفة).

وبهدف عمل مزيد من التجارب حول تأثير عدد جمل التدريب على دقة التشكيل، فقد تم اختبار كفاءة أداء النظام بتدريبه بعشرين ألف جملة في خمس تجارب وبحوالي أربع وعشرين ألف جملة في تجربة سادسة. وكانت النتائج كما هي معروضة في الجدول: ١. ويلاحظ هنا أن نسبة صحة التشكيل ضلت على ما هي عليه، وهذا يؤكد أن عدد جمل التدريب لا تعطي اختلافا ذا قيمة عندما يكون عددها بين عشرة آلاف جملة وأربع وعشرين جملة.

الرقم	جمل التدريب	جمل الاختبار	نسبة الصحة
١	٢٠٠٠	٤٠٠	٧٢.٩٠
٢	٢٠٠٠	٨٠٠	٧٢.٩٤
٣	٢٠٠٠	١٢٠٠	٧٢.٧٨
٤	٢٠٠٠	١٦٠٠	٧٢.٦٧
٥	٢٠٠٠	٤٠٠	٧٢.٩٣
٦	٢٤٢٧٤	٢٤٢٧٤	٧٢.٧٣

الجدول: ١. نسبة الصحة للنظام uni-model، الأرقام العريضة تعني أن

نصوص الاختبار تختلف عن نصوص التدريب.

ولمحاولة رفع كفاءة النظام، قمنا بإضافة بعض القواعد اللغوية على الاحتمالات المتقطعة (الاحتمالات المستخدمة من قبل الـ HTK لتوليد علامات التشكيل). أما بالنسبة للحروف التي يمكن تطبيق هذه القواعد عليها فيمكن تقسيمها إلى عدة أقسام وهي:

١. الحروف التي يجب أن تشكل:

كل أرقام الحروف التي في الملحق: ٢. ماعدا (١، ٤، ١٠، ١١، ١٢، ١٣، ١٤، ١٥، ٢٨، ٢٩، ١٠١، ١٠٢، ١٠٣، ١٠٤، ١٠٥، ١٠٧، ١٠٨، ١٠٩، ١١٠). هذه الحروف يجب أن تكون قيمتها ٠ في الموديل non.

٢. الحروف التي يجب ألا تشكل:

الحروف التي أرقامها (١١، ١٢، ١٣، ١٤، ١٥، ١٠٩، ١٠٥). هذه الحروف يجب أن تكون قيمتها "١" في الموديل non أي أنها لا تشكل (ف non هنا تعني بدون تشكيل) و "٠" في كل الموديل الأخرى.

٣. الحروف التي تأتي دائماً مع حالة الصمت SIL:

هنا يوجد حرف واحد فقط وهو الحرف رقم (١١٠). هذا الحرف يجب أن تكون قيمته "٠" في كل الموديل ماعدا الموديل SIL حيث يجب أن تكون قيمته فيه تساوي "١".

٤. الأحرف التي لا يمكن أن تظهر في بعض المواضع:

هناك عدة حروف لا يمكن أن تأتي في مواضع معينة في الكلمة العربية. مثلاً حرف: "ؤ" لا يمكن أن يأتي في بداية الكلمة. أرقام هذه الحروف هي (١، ٤، ٧، ٢٨، ٢٩، ١٠٣، ١٠٤). وهذه الحروف يجب أن تكون قيمتها ٠ في كل الموديل بلا استثناء.

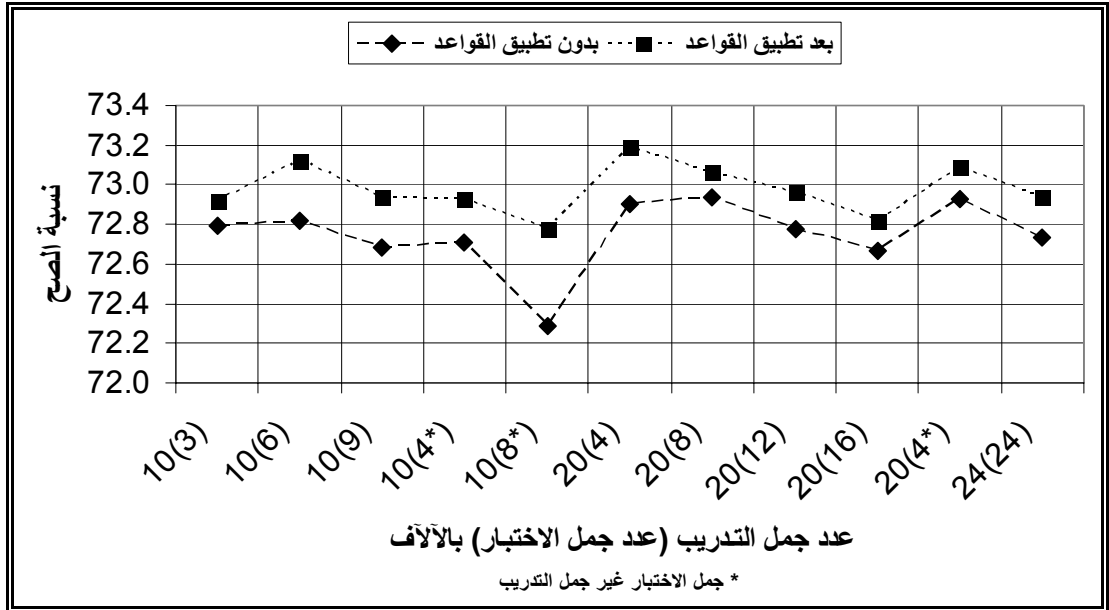
٥. الحروف التي لا تأتي مع بعض علامات التشكيل:

مثلاً علامة التشكيل msk (شدة مع كسرة) لا يمكن أن تأتي مع الحرف رقم ١٠ (الألف في بداية الكلمة)، وفي هذه الحالة نضع القيمة العاشرة في الموديل msk تساوي "٠". وهذه العملية تطبق على كل الحالات المماثلة وذلك حسب الجدول الموضح في الملحق: ١.

وقد أدت إضافة القواعد اللغوية المشار إليها آنفا إلى تحسن طفيف في نسبة التشكيل الصحيح كما هو موضح في النتائج المعروضة في الجدول: ٢. والمقارنة بين نتائج الحالتين كما هو مبين في الشكل: ٣.

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	١٠٠٠٠	٣٠٠٠	٧٢.٩٢
٢	١٠٠٠٠	٦٠٠٠	٧٣.١٢
٣	١٠٠٠٠	٩٠٠٠	٧٢.٩٤
٣	١٠٠٠٠	٤٠٠٠	٧٢.٩٣
٤	١٠٠٠٠	٨٠٠٠	٧٢.٧٨
٥	٢٠٠٠٠	٤٠٠٠	٧٣.١٩
٦	٢٠٠٠٠	٨٠٠٠	٧٣.٠٦
٧	٢٠٠٠٠	١٢٠٠٠	٧٢.٩٦
٨	٢٠٠٠٠	١٦٠٠٠	٧٢.٨٢
٩	٢٠٠٠٠	٤٠٠٠	٧٣.٠٩
١١	٢٤٢٨٨	٢٤٢٨٨	٧٢.٩٤

الجدول: ٢. النتائج بعد تطبيق القواعد النحوية والصرفية، الأرقام العريضة تعني أن نصوص الاختبار تختلف عن نصوص التدريب.



الشكل: ٣. رسم بياني يوضح مقارنة بين نسب الصحة لنظام uni-model (ذخيرة المشروع).

ولمزيد من التجارب على النظام، قمنا باستخدام قاعدة بيانات الأسماء العربية المشار إليها سابقا. فدرينا النظام عليها وكانت أفضل نتيجة تم التوصل لها هي ٧٨.٧٨٪، لكننا قمنا بعد ذلك ببعض التعديلات في طريقة تشكيل هذه الأسماء ومن ثم تدريب النظام واختباره بهذه الذخيرة في ثلاث تجارب نتائجها موضحة في الجدول: ٣. والسبب في تحسن أداء النظام عند استخدام قاعدة البيانات هذه خلوها من الحالات الإعرابية أي أن تشكيل آخر الاسم ثابتا دائما (وهو الحرف الأخير للاسم بدون حركة والحرف الأخير للكلمة الأولى في الأسماء المركبة مضموم)، إضافة إلى التطابق الكبير بين الأسماء كورود "عبد" ذات التشكيل الثابت مع أسماء الله الحسنى "الرحمن والهادي والعزيز..."

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	٥٥٠٠٠	٥٥٠٠٠	٨٠.٠٥
٢	٥٥٠٠٠	٢٠٠٠٠	٨٠.٥٧
٣	٥٥٠٠٠	٤٠٠٠٠	٨٠.٠٧

جدول ٣: نسبة الصحة للنظام uni-model (قاعدة بيانات الأسماء العربية).

ثم قام الفريق بتدريب النظام على نصوص القرآن الكريم ومن ثم اختبار كفاءته من خلال ثلاث تجارب كانت نتائجها على نحو ما هو معروض في الجدول: ٤.

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	٥٣٠٨	٥٣٠٨	٧٠.٩٢
٢	٥٣٠٨	٢٠٠٠	٧٠.٩١
٣	٥٣٠٨	٤٠٠٠	٧٠.٩٣

الجدول ٤: نتائج تدريب النظام بنصوص القرآن الكريم.

استخدام أنموذج التسلسل الثلاثي للحروف tri-model:

بعد الانتهاء من التجارب المختلفة على النظام باستخدام التسلسل المفرد للحروف uni-model قمنا بالتحول إلى خيار أنموذج التسلسل الثلاثي للحروف tri-model ، حيث تم تدريب النظام واختباره بعدة نصوص بما فيها ذخيرة المشروع وقاعدة بيانات الأسماء ونصوص القرآن الكريم. ويبين الجدول: ٥. أن النتائج باستخدام التسلسل الثلاثي أفضل بكثير من تلك الناتجة عن تطبيق التسلسل المفرد. حيث ارتفعت النسبة الصحيحة إلى ما يقارب ١٠٪. والسبب في ذلك أن التسلسل الثلاثي يأخذ بتأثير الحروف المتجاورة على تشكيل كل حرف فيها، ومن المعلوم أنه كلما ارتفع عدد الحروف في التسلسل كلما كان ذلك متطابقا مع العبارة المراد تشكيل حروفها إلا أن التسلسل الثلاثي هو أعلى حد تسمح به أدوات

ماركوف الخفية لسبب منطقي وهو أنه كلما ارتفع عدد تسلسل الحروف كلما كان ذلك على حساب سرعة أداء النظام.

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	٥٠٠٠	٥٠٠٠	٨٣.٠٦
٢	١٠٠٠٠	١٠٠٠٠	٨٣.٢٣
٣	١٥٠٠٠	١٥٠٠٠	٨٣.٥٣
٤	٢٤٢٨٨	٢٠٠٠٠	٨٠.١٩
٥	٢٤٢٨٨	٥٠٠٠	٧٨.٢٦
٦	٢٤٢٨٨	١٠٠٠٠	٧٩.٠٣
٧	٢٤٢٨٨	١٥٠٠٠	٧٩.٦٧
٨	٢٤٢٨٨	٢٤٢٨٨	٨٣.٤٤
٩	٥٠٠٠	٣٠٠٠	٨٢.٢٩
١٠	١٠٠٠٠	٣٠٠٠	٨٣.٢٧
١١	١٥٠٠٠	٣٠٠٠	٨٣.٠٣
١٢	٢٠٠٠٠	٣٠٠٠	٨٢.٤٦
١٣	٥٠٠٠	٥٠٠٠	٨٣.٣٥
١٤	١٠٠٠٠	٥٠٠٠	٨٣.٣٥
١٥	١٥٠٠٠	٥٠٠٠	٨٣.١٥

الجدول: ٥. نسبة الصحة للنظام tri-model، الأرقام العريضة تعني أن نصوص الاختبار تختلف عن نصوص التدريب.

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	٥٥٠٠٠	٥٥٠٠٠	٨٨.٧٩
٢	٥٥٠٠٠	٢٠٠٠٠	٨٩.٨١
٣	٥٥٠٠٠	٤٠٠٠٠	٨٩.٠٩

الجدول: ٦. نسبة الصحة للنظام tri-model. (قاعدة بيانات الأسماء العربية).

وبالنظر إلى النصوص الناتجة من النظام، وجدنا أن نسبة كبيرة من الأخطاء تظهر على الحرف الأخير من الكلمة، ولهذا السبب قمنا بعمل بعض التعديلات في طريقة عمل الـ HTK بحيث يقوم النظام بتشكيل النصوص العربية مع تجاهل الحرف الأخير من كل كلمة (على مبدأ "سكن تسلم"). ونتج

عن ذلك ارتفاع نسبة التشكيل الصحيح لتتجاوز ٨٩،١٤ بينما كانت ٨٣،٤٤ عندما يكون تشكيل الحرف الأخير أحد متطلبات التشكيل.

وبالإضافة إلى استخدام ذغيرة المشروع، قمنا بتدريب النظام واختباره بذخيرة الأسماء العربية وكانت أفضل نتيجة تم التوصل لها هي ٨٩،٣٠. ولكن بعد تطبيق بعض التعديلات في طريقة تشكيل هذه الأسماء (الجدول: ٠٦).

أيضاً قمنا بتدريب النظام على نصوص القرآن الكريم ومن ثم اختبارها من خلال ثلاث تجارب ظهرت نتائجها في الجدول: ٠٧.

الرقم	نصوص التدريب	نصوص الاختبار	نسبة الصحة
١	٥٣٠٨	٥٣٠٨	٨٣،٨١
٢	٥٣٠٨	٢٠٠٠	٨٣،٣٧
٣	٥٣٠٨	٤٠٠٠	٨٣،٧٢

الجدول: ٠٧. نسبة الصحة للنظام tri-model. (نصوص القرآن الكريم).

وبناء على التجارب والنتائج السابقة، فقد تم بناء نظام حاسوبي متكامل لتشكيل النص العربي اعتماداً على الـ HTK (الشكل: ٤؛ والشكل: ٥). و هذا البرنامج يتيح للمستخدم أن يكتب النص العربي المراد تشكيله أو أن يختار نصاً جاهزاً. ويمر هذا البرنامج عبر ثلاث برمجيات هي:

البرمجيات الأولى: تقوم بتحويل النص العربي إلى نصوص ثنائية بصيغة محددة ذكرت آنفاً ليستطيع الـ HTK التعامل معها،

البرمجيات الثانية: برمجيات Hvite هي إحدى البرمجيات الجاهزة في حزمة الـ HTK وتقوم بتوليد علامات التشكيل،

البرمجيات الثالثة: تقوم بعملية معاكسة للبرمجيات الأولى، حيث تقوم بقراءة مخرجات الـ HTK ومن ثم تحويلها إلى نص عربي مشكل. ويتيح هذا النظام للمستخدم خيارين للتشكيل هي:

أ. تشكيل كامل حروف الكلمة:

من الخيارات المتاحة، تستطيع أن تجعل البرنامج يقوم بتشكيل النص العربي كاملاً بلا أي استثناء للحرف الأخير من الكلمة. هذا الخيار يعتمد على أنموذج tri-model. لذلك فإن نسبة الصحة للنصوص الناتجة من هذا الخيار تتراوح ما بين ٨٣٪ إلى ٨٥٪.

ب. تجاهل الحرف الأخير:

كما تمت الإشارة إليه سابقاً، فإن نسبة كبيرة من الأخطاء في التشكيل تظهر على الحرف الأخير من الكلمة، لذلك قمنا بوضع هذا الخيار لرفع كفاءة النظام. وهذا الخيار يعتمد على أنموذج tri-model لكن بإضافة برنامج مبسط يفرض على الـ HTK تجاهل الحرف الأخير. ونسبة النصوص المشكولة من هذه الطريقة تتراوح ما بين ٨٨٪ إلى ٩١٪.



الشكل: ٤. واجهة استخدام نظام التشكيل.



الشكل: ٥. واجهة النظام بخيارات أخرى.

٣.٢.٢. نظام تشكيل يعتمد على Viterbi Algorithm (النظام الثاني):

يعتمد مصممو الخوارزميات في الغالب على استخدام خوارزميات فيتربي Viterbi Algorithm على مستوى الحروف ولكننا قمنا بتطبيق هذه الخوارزمية على مستوى الكلمات. وقبل أن نبدأ بتصميم النظام كان علينا أن نقوم بكتابة ستة برمجيات مبسطة لإعداد قاعدة البيانات اللازمة لهذه الخوارزمية حتى تقوم بعملها على الوجه المطلوب.

وبعد الانتهاء من تصميم النظام وكتابة جميع البرمجيات المتعلقة به، قمنا باختبار كفاءته وذلك بإدخال نص عربي مكون من ألف كلمة عربية و عندما قمنا بمقارنة الملف الأصلي بالملف الناتج وجدنا أن هناك ثمانية أخطاء فقط في عملية التشكيل كل خطأ من هذه الأخطاء ظهر في كلمة مستقلة.

قيود البرنامج:

- يتطلب هذا البرنامج أن تكون جميع الكلمات المدخلة إليه موجودة في قاعدة بيانات البرنامج حتى يتم تشكيلها. ومن ثم فإن هذا البرنامج لن يقوم بتشكيل أية كلمة غير موجودة في قاعدة بياناته.
- يخرج النص المشكل بدون علامة تشكيل الحرف الأخير.

وقد كتبت ورقتان علميتان منشورتان تشرح تفاصيل هذا النظام (Elshafei, Almuhtasib and Alghamdi 2006a, 2006b).

٣.٢.٣ نظام التشكيل الآلي المستقل (النظام الثالث):

إنطلاقاً من سياسة معهد بحوث الحاسب والإلكترونيات بأهمية توفر البرمجيات المفتوحة المصدر، إضافة إلى حرص الفريق العلمي في المشروع على أن تكون جميع البرمجيات التي ينتجها مفتوحة المصدر، فقد كانت هناك عدة محاولات لبناء نظام مستقل عن أنموذج ماركوف الخفي HMM. إلا أن الفريق واجهته معضلات عدة كان منها تشغيل النظام الجديد دون عيوب bugs وفي حالة العمل فإن النسبة الصحيحة للتشكيل تقع حول ٥٠%. وكان من أسباب ذلك أن الفريق حاول محاكاة أدوات ماركوف الخفية HTK ولكن باستخدام لغات برمجة كاسي بلس بلس C++.

وللخروج من هذه الدائرة كان على الفريق التفكير في طريقة مختلفة تماماً. فكان نظام التشكيل الآلي المستقل. وتقوم فكرة النظام المستقل على أن لتشكيل النص العربي نسق معين ينسجم مع الحروف ويتبع تسلسلها، بمعنى أن للتشكيل تسلسل شبه منتظم يتفق مع تسلسل الحروف، وهذا بمنأى عن القواعد اللغوية من صرف ونحو وخلافه. ولو استطعنا تتبع جميع تسلسلات الحروف لعرفنا تسلسل الحركات. مثال ذلك تسلسل الحروف التالية:

$$(١) \quad \# \text{ ي ت ت ب ع } \# \text{ ا ل ن ا س } \# \text{ م ا } \# \text{ ي س ت ج د } \# \text{ م ن } \# \text{ ا ح د ا ث } \#$$

حيث # بداية العبارة أو المسافة بين الكلمات.

فعند ورود التسلسل (١) فإنه من الممكن معرفة التشكيل الذي يتبع كل حرف فيه، بغض النظر عن معنى العبارة أو قواعدها النحوية أو الصرفية، فنحن هنا نتعامل مع التسلسل المجرد للحروف دون إرتباطها اللغوي. التسلسل (١) له التسلسل التشكيلي الثابت التالي:

$$(٢) \quad \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \# \text{ _ _ _ _ } \#$$

حيث: " _ " يمثل موقع الحرف في تسلسل الحروف (١).

فلو ورد تسلسل الحروف (١) وطلب تشكيله فإن التشكيل (٢) هو التشكيل الصحيح له.

هذا يعني أنه (نظريا) لو أمكن حصر جميع حالات تسلسل الحروف في اللغة العربية لأمكن معرفة تشكيل جميع حروف العبارة. إلا أن المعضلة تكمن في أمرين: الأمر الأول، أن تسلسل الحروف متعدد لدرجة أنه بعيدا عن الحصر، ففي التسلسل (١) يمكن إضافة أو حذف أو تعديل أي من الكلمات أو أجزاء من الكلمات المكونة لهذا التسلسل، فيتغير تسلسل الحروف ومن ثم يتغير تشكيلها. الأمر الثاني، أنه ليس هناك تشكيلا لتسلسلات الحروف جميعها. فالمتمون المشكلة محدودة جدا. وقد يمكن تجاوز هاتين المعضلتين في المستقبل مع التقدم في سرعة معالجة النصوص وتخزين البيانات وإسهامات اللغويين في تشكيل النصوص العربية.

ولأن خيار تشكيل كامل العبارة غير ممكن في الوقت الراهن على الأقل، فإنه لا بد من البحث عن خيارات بديلة. أحد هذه الخيارات هو استخدام تسلسل لعدد حروف أقل. ولكن ما هو عدد التسلسل المناسب؟ ولتوضيح ذلك فإننا سنأخذ أقل تسلسل ممكن وهو واحد. فالملاحق: ٣ يبين تسلسل الحروف العربية وتكرار ورودها مع الحركات المختلفة في المتن الذي قام فريق المشروع بتشكيل نصوصه. فلو أخذنا أعلى تكرار، بمعنى آخر، أعلى احتمالية لتشكيل كل حرف لكان هناك خيارا واحدا للتشكيل وغالبا ما تكون الفتحة. أي أن تشكيل العبارة السابقة في تسلسل الحروف (١) هو:

يَتَّبِعُ النَّاسُ مَا يَسْتَجِدُّ مِنْ أَحْدَاثٍ

ويلاحظ هنا كيف أن الفتحة هي المهيمنة على الخيارات الأخرى لكونها تستحوذ على أعلى تكرار في تشكيل معظم الحروف. ولأن الحرف الواحد هو أقل تسلسل ممكن وتسلسل جميع حروف العبارة هو أعلى تسلسل للحروف، وكلاهما غير ممكن للأسباب المذكورة آنفا، كان لا بد من إختيار تسلسل ممكن تنفيذه برمجيا ويعطي أعلى درجة ممكنة من الدقة في التشكيل. فلو أخذنا خيار تسلسل حرفين bi-gram لكان عدد احتمالات ورود تسلسل كل حالتين هو:

$$1225 = {}^2(35)$$

حيث أن عدد حروف الكتابة العربية ٣٤ حرفا هي: ء أ إ ئ ا آ ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع غ ف ق ك ل م ن و ه ة ي . ويضاف لها الفراغ بين الكلمات فتصبح ٣٥.

إلا أننا وجدنا أن نسبة الخطأ باستخدام هذه الطريقة مرتفعة وليست أفضل من الأولى بدرجة كبيرة. لهذا أخذنا الحالة التي تليها وهي تسلسل الثلاثة حروف tri-gram فوجدنا أن نسبة الخطأ لا تزال مرتفعة فأخذنا الحالة التالية وهي التسلسل الرباعي quad-gram. فكانت النسبة مشجعة كثيرا خاصة وأننا لم نضف قواعد خاصة بتشكيل بعض الحروف كما فعلنا مع النظام الذي يعتمد على أدوات ماركوف الخفية (النظام الأول). والتسلسل الرباعي للحروف يولد عددا كبيرا من التسلسلات :

$$1,000,625 = {}^4(35)$$

وهذا عدد كبير لحالات التسلسل الرباعي للحروف العربية. إلا أنه كما هو معروف فإنه ليست جميع هذه التسلسلات ممكنة في اللغة. فعلى سبيل المثال، لا تسمح الكتابة العربية بالتسلسلات التالية:

ة ع ف

ك ي ل م

ح ج ع غ

ومع الأبقاء على فرضية تسلسل جميع الحروف العربية، فإنه عند إضافة احتماليات التشكيل إلى احتماليات تسلسل الحروف فإن العدد يتضاعف بدرجة كبيرة (١٤×٣٥). حيث ٣٥ عدد الحروف إضافة إلى الفراغ و ١٤ هو عدد الحركات () فيصبح إجمالي الاحتمالات: ١١٧،٦٤٩،٠٠٠ وهذا الرقم كبير جدا ومن الصعوبة بمكان دمجها في نظام للتشكيل. لهذا كان لا بد من البحث عن طريقة لاستخراج أقل عدد ممكن من التسلسل الرباعي بحيث لا يؤثر على دقة التشكيل. فكان الحل هو استخلاص التسلسل الرباعي للحروف ذات أعلى احتمالية تشكيل وهي التسلسلات الرباعية الفاعلة في التشكيل فكان عددها ٦٨،٣٧٨ تسلسلا فقط (عينة في الجدول: ٨). هذا العدد من التسلسلات الرباعية للحروف استخلصت من الذخيرة اللغوية التي أعدها فريق المشروع.

الاحتمالية	التشكيل ٥		الحرف ٤	التشكيل ٣		الحرف ٢	التشكيل ١		الحرف ١	التشكيل	الحرف
٠.١٨	٠		م	٠	٠	#	٠		د	٠	ج
١.٠٠	٠		هـ	٠		د			د	٠	ج
١.٠٠	٠		ء	٠	٠	ا	٠		د	٠	ج
٠.٩٢	٠	٠	#	٠		ر	٠		د	٠	ج

الجدول: ٨. عينة من التسلسل الرباعي للحروف (# = الفراغ بين الكلمات أو في نهايتي الجملة)

ومن ثم وضعت الفرضية التالية لتشكيل النص العربي:

يمكن تشكيل نص عربي بتمرير احتمالية التشكيل للتسلسل الرباعي
للحروف بداية بأول حرف في العبارة وانتهاء بأخر حرف والأخذ بأعلى
احتمالية تشكيل لكل حرف.

ولاختبار هذه الفرضية فقد تم بناء نظام حاسوبي بلغة جافا JAVA (الشكل: ٦) يتكون كما هو موضح
في الشكل: ٧ (المخطط التجميعي للنظام) من وحدتين أساسيتين:

الوحدة الأولى: وحدة التحليل والمقارنة،

ويكون مدخل هذه الوحدة النص المطلوب تشكيله وقائمة التسلسلات الرباعية للحروف المشكلة، ويتم
في هذه العملية ما يلي:

١. تحليل النص غير المشكل إلى سلسلة من التسلسلات الرباعية للحروف المكونة له.
٢. استدعاء التسلسلات الرباعية للحروف المشكلة والتي تقابل التسلسلات الرباعية للحروف الغير
مشكلة والمطلوب تشكيلها، وفي حالة عدم وجود تسلسل مقابل في قائمة التسلسل الرباعي

للحروف المشكلة فإن النظام يضع تسلسلا رباعيا افتراضيا غير مشكل يقابل التسلسل الرباعي للحروف غير المشكلة وهذا فقط لاستمرارية عمل البرمجيات دون إعاقة. وينتج عن هذه الوحدة قائمة من التسلسلات الرباعية لحروف النص المراد تشكيلا مع تشاكيها المختلفة واحتمالية ورود كل تسلسل. الوحدة الثانية: اختيار التشكيل الأعلى احتمالية، إذ يصبح منتج الوحدة الأولى مدخلا للوحدة الثانية، فيصبح لكل حرف أربع احتمالات لتشكيله كحد أعلى. ثم تجمع الاحتمالات الموحدة وتدمج في احتمال واحد. بعد ذلك يتم اختيار التشكيل صاحب أعلى احتمالية لكل حرف بداية بالحرف الأول ثم الذي يليه حتى ينتهي النص. ولعرض الخطوات السابقة في مثال لنص يراد تشكيله، فقدتم إدخال العبارة "المشكل الآلي للحرف العربي" فقام النظام بعمل التالي:

١. تحليل مكونات العبارة إلى تسلسلات رباعية الحروف وهي:

ا ل م

ا ل م ش

ل م ش ك

م ش ك ل

ش ك ل #

ك ل # ا

ل # ا ل

ا ل آ

ا ل آل

ل آل ي

آ ل ي #

ل ي # ل

ي # ل ل

ل ل ح

ل ل ح ر

ل ح ر ف

ح ر ف #

ر ف # ا

ف # ا ل

ا ل ع

ا ل ع ر

ل عرب

ع ر بي

ر بي #

حيث ترمز # للفراغ space بين الكلمات أو نهايتي العبارة.

٢. استخراج التسلسل الرباعي للحروف من قائمة الحروف المشكلة والمطابقة لقائمة التسلسل الرباعي

للحروف غير المشكلة، وهي كالتالي مع ذكر احتمالية صحتها:

٠.٤٩	٠	م	٠	ل	٠	٠	ا	٠	٠	#
٠.٣٤	٠	ش	٠	م	٠	ل	٠	٠	ا	
٠.٨٩	٠	ك	٠	ش	٠	م	٠	ل	٠	
٠.٩٩	٠	ل	٠	ك	٠	ش	٠	م	٠	
٠.٤١	٠	#	٠	ل	٠	ك	٠	ش	٠	
٠.٣٧	٠	ا	٠	#	٠	ل	٠	ك	٠	
٠.٣٣	٠	ل	٠	ا	٠	#	٠	ل	٠	
٠.٩٢	٠	آ	٠	ل	٠	ا	٠	#	٠	
٠.٥٢	٠	ل	٠	آ	٠	ل	٠	ا	٠	
٠.٥١	٠	ي	٠	ل	٠	آ	٠	ل	٠	
٠.٦٩	٠	#	٠	ي	٠	ل	٠	آ	٠	
٠.٢٧	٠	ل	٠	#	٠	ي	٠	ل	٠	
٠.٣٤	٠	ل	٠	ل	٠	#	٠	ي	٠	
٠.٦١	٠	ح	٠	ل	٠	ل	٠	#	٠	
٠.٦١	٠	ر	٠	ح	٠	ل	٠	ل	٠	
٠.٦٦	٠	ف	٠	ر	٠	ح	٠	ل	٠	
٠.٣٤	٠	#	٠	ف	٠	ر	٠	ح	٠	
٠.١٨	٠	ا	٠	#	٠	ف	٠	ر	٠	
٠.٣٨	٠	ل	٠	ا	٠	#	٠	ف	٠	
٠.٧٣	٠	ع	٠	ل	٠	ا	٠	#	٠	
٠.٧٢	٠	ر	٠	ع	٠	ل	٠	ا	٠	
٠.٩٣	٠	ب	٠	ر	٠	ع	٠	ل	٠	
٠.٧٠	٠	ي	٠	ب	٠	ر	٠	ع	٠	
٠.٣٦	٠	#	٠	ي	٠	ب	٠	ر	٠	

استنباط الاحتمالية الأعلى لتشكيل كل حرف: وسنبدأ بالحرف الأول وهو "ا" حيث ورد في التسلسل الأول والثاني فقط وكانت علامة التشكيل التي تتبعه "٠" في كلا الحالتين، هذا يعني أن هذا الحرف

لن يضع النظام عليه علامة تشكييل. ثم نأتي للحرف الذي يليه وهو "ل" الذي ورد في التسلسل الأول والثاني والثالث، وكانت علامة التشكييل في جميع هذه الحالات "°" فسيضع النظام علامة السكون على هذا الحرف. والحرف الثالث في النص المطلوب تشكييله هو "م" وقد ورد في التسلسل الأول والثاني والثالث والرابع، وكانت علامة التشكييل المصاحبة له فيها جميعا هي "´" فسيكون تشكييله كذلك. والحرف الرابع في النص غير المشكل هو "ش" الذي ورد في التسلسل الثاني والثالث والرابع والخامس، وكان تشكييله في التسلسل الثاني والثالث والرابع "°" بينما في التسلسل الخامس كان تشكييله "´"، هذا يعني أن النظام سيجمع الاحتمالات الواردة في التسلسلات الثلاثة الأولى لأن لها التشكييل نفسه = ٠.٣٤+٠.٨٩+٠.٩٩= ٢.٢٠ وسيأخذ بعلامة التشكييل "°" حيث احتمالاتها هي الأعلى (٢.٢٠) للسكون و ٠.٤١ للفتحة). وهكذا مع بقية الحروف حتى ينتهي النص.

وقد كان تشكييله للعبارة السابقة هو "المُشكَلُ الآليُّ لِلْحَرْفِ العَرَبِيِّ". ويلاحظ هنا ارتفاع نسبة التشكييل الصحيح رغم أن النظام يتعامل مع تسلسل الحروف فقط ولم تضاف قواعد صرفية أو نحوية أو دلالية لعملية التشكييل فيه. لهذا يلاحظ أن تشكييله للكلمة الأولى لا يتناسب ومعناها هنا لأن ذلك يتطلب إضافة عامل الدلالة وإلا فإن التشكييل صحيح من حيث أن الكلمة تقع ضمن كلمات المعجم العربي. كما أن قواعد النحو والصرف مهمة لتشكييل أواخر الكلمات ولو أن النظام قام بتشكييل أواخر الكلمتين الثانية والثالثة تشكيلا صحيحا رغم أنه لم يبرمج على قواعد النحو في التشكييل وإنما اعتمد على التسلسل الرباعي للحروف.

```

sequence_count = -1;

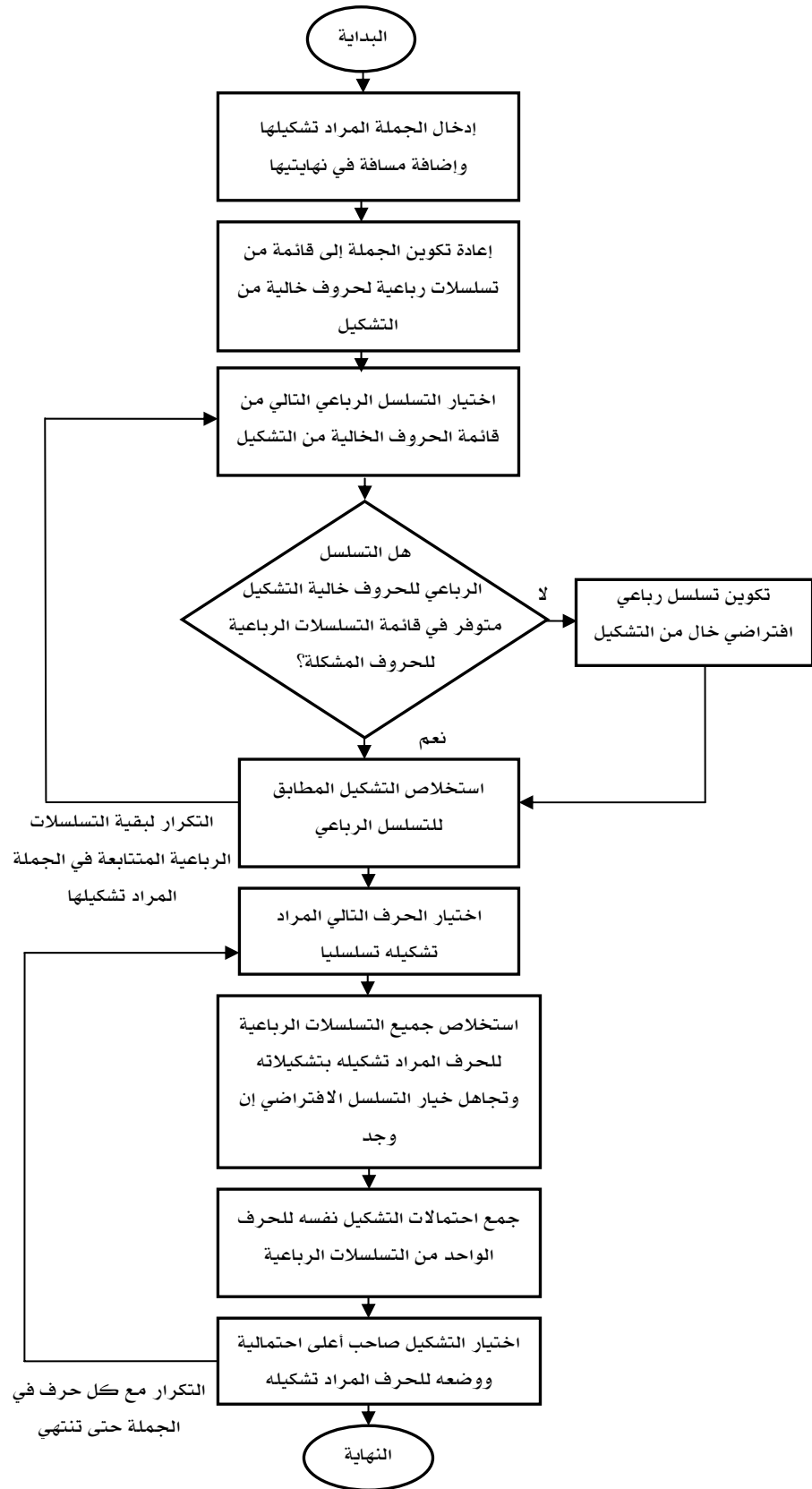
for(ind = (numSeq-1); ind >=0; ind--) {
    sequence_count++;
    alphabetTri[sequence_count] = (triAlphabets)dataTriCom.get(index-ind); // sequence 1
    if((alphabetTri[sequence_count].getAlphabet(ind+1)).getAlphabet() == letter) {
        selCount++;
        alpha[selCount] = alphabetTri[sequence_count].getAlphabet(ind+1);
        diacPercent[selCount] = alphabetTri[sequence_count].
            getFrequency();
    }
}

for(ind = 0; ind <= selCount; ind++) {
    diac1 = alpha[ind].getFirstDiacritic();
    diac2 = alpha[ind].getSecondDiacritic();
    //System.out.print("Diac1 = " + diac1 + "and" + "Diac2 = " + diac2);
    comDiacCode[ind] = diacCode(diac1, diac2);
}

extractedCode = findDiacritic(comDiacCode, diacPercent, selCount+1);
extractedDiac = codeDiac(extractedCode);

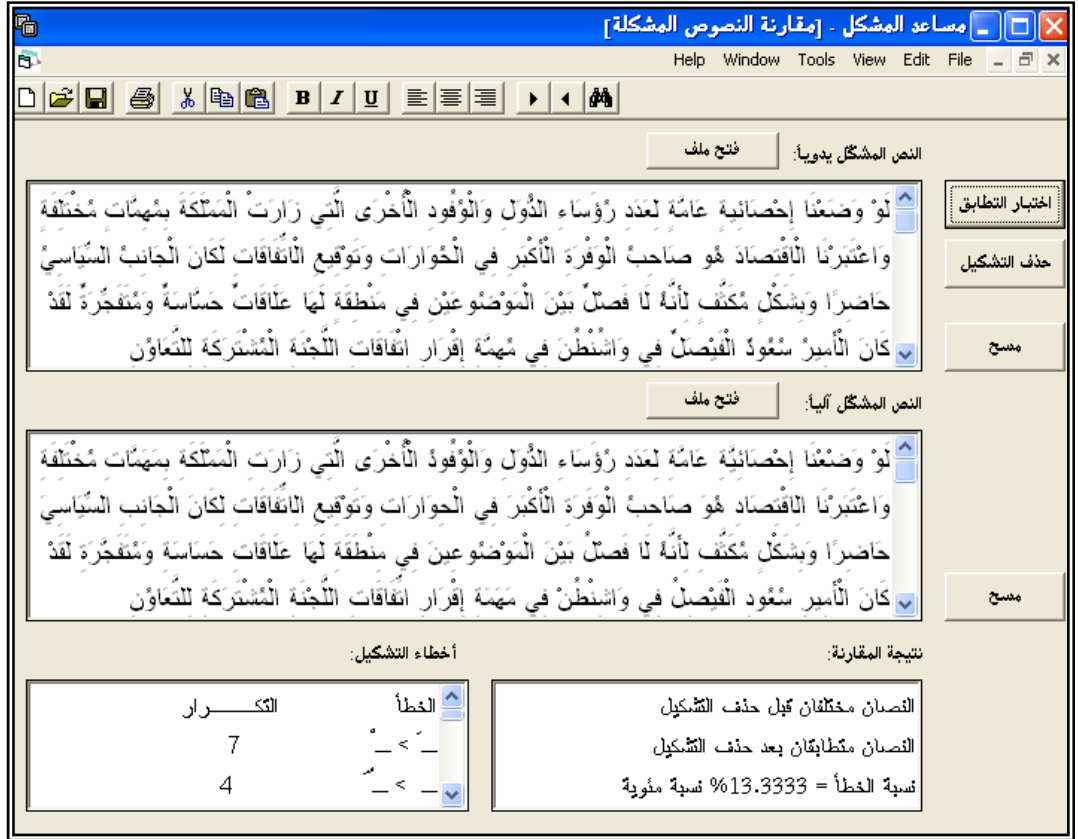
```

الشكل: ٦. مقطع للشفرة code التي كتبت بها برمجيات نظام التشكييل الآلي المستقل.



الشكل: ٧. المخطط التجميعي لنظام المشكل الآلي المستقل.

ولإختبار قدرة نظام تشكييل الحرف العربي على دقة التشكييل ومقارنة ذلك بالنظام الذي يقوم على أدوات ماركوف الخفية فقد قام المهندس عبد الله الزامل بكتابة برمجيات للمقارنة (الشكل: ٨). ثم أخذ نص حديث وهو إفتتاحية جريدة الرياض اليومية (الاثنين ٢٤ ربيع الآخر ١٤٢٧هـ - ٢٢ مايو ٢٠٠٦م - العدد ١٣٨٤٦) ثم شكل يدويا ثم طلب من نظام التشكييل الذي يعمل بأدوات ماركوف الخفية (الأنموذج الثلاثي tri-model) ونظام التشكييل المستقل بتشكييله فكانت النتيجة: ٧٤٪ للأول و ٨٧٪ للآخر.



الشكل: ٨. واجهة مقارنة النصوص المشكلة آليا بالنصوص المشكلة يدويا.

ويلاحظ انخفاض نسبة المشكل الآلي بأدوات ماركوف الخفية هنا عن النسب التي أشير لها سابقا تحت نظام تشكييل يعتمد على أدوات ماركوف الخفية (الأنموذج الثلاثي tri-model) والسبب في ذلك أن النسب التي توردها أدوات ماركوف سواء في هذا التقرير أو في الأوراق العلمية الأخرى المنشورة والمشار لها تحت قائمة المراجع في هذا التقرير هي نسب غير دقيقة. ذلك لأننا قمنا بمقارنة نسب التشكييل الصحيح التي تولدها أدوات ماركوف الخفية والنسب التي تقوم بحسابها برمجيات عبد الله الزامل فوجدنا أن هناك ما معدله ١٠٪ زيادة في نسب أدوات ماركوف الخفية، هذا يعني أن أدوات ماركوف الخفية تعطي نسبة الصحة في التشكييل ١٠٪ أعلى من الواقع. والحقيقة أن السبب ليس لأن أدوات ماركوف الخفية تبالغ في أدائها ولكن لسبب تقني بحت يتعلق بطريقة عمل هذه الأدوات.

فالبيانات التي تأخذها أدوات ماركوف دائماً في الحسبان عند وضع نسبة التشكيل الصحيح تشمل جميع الراصدات بما فيها بداية العبارة ونهايتها والفراغ بين الكلمات والتي يرمز لها حسب الملحق: ٢ ب (١١٠ و ١٠٩ على التوالي) هذه الراصدات لها دائماً الواسمات نفسها وهي non. وفي هذه الحالة فإن الراصدات تشكّل في واقع الأمر ١٠٪ تقريباً من إجمالي عدد الراصدات في النص. وحيث أن تشكيلها ثابت وهو بدون تشكيل فإن النظام يظليها لنسبة الصحة في التشكيل الذي ينتجه.

مميزات نظام التشكيل الآلي المستقل:

١. الإستقلالية عن البرمجيات المقيدة جزئياً أو كلياً. فالنظام الجديد مستقل تماماً ويمكن برمجته بأية لغة برمجه والتصرف فيه بحثياً وتجارياً.
٢. يتعامل مع الحرف العربي مباشرة. وهذا بعكس النظم الأخرى التي تقوم أولاً بتحويل الحرف العربي إلى روماني ثم إختيار علامات التشكيل المناسبة وبعد ذلك يحول الحرف الروماني إلى حرف عربي مشكل.
٣. السرعة في الأداء: فمعدل فترة تشكيل الكلمة الواحدة هو ١,٧ مليثانية (١,٧/١٠٠٠ ثانية) بينما معدل فترة تشكيل الكلمة في نظام التشكيل بأدوات ماركوف الخفية هي ٨٤,٢ مليثانية (٨٤,٢/١٠٠٠ ثانية) على الحاسب نفسه وفي الظروف نفسها. أي أنه يستغرق ٥٠/١ من الوقت الذي يستغرقه نظام التشكيل بأدوات ماركوف الخفية عند وضع علامات التشكيل على نص غير مشكل.
٤. صغر الحجم مقارنة بالنظم الأخرى. فحجم نظام التشكيل المستقل لا يتجاوز بدون ضغط ٣ ميغابايت وبعد ضغطه أقل من نصف ميغابايت. بينما حجم نظام التشكيل الذي يعتمد على أدوات ماركوف الخفية ٢٦ ميغابايت.
٥. ارتفاع نسبة التشكيل الصحيح. فمعدل نسبة التشكيل الصحيح فيه لجميع حروف العبارة بما فيها آخر الكلمات هو ٨٧٪ من إجمالي عدد الحروف بينما نسبة التشكيل الصحيح في نظام التشكيل بأدوات ماركوف الخفية هي ٧٤٪.
٦. النظام لا يستخدم أية قواعد أو قوانين نحوية أو صرفية، فهو يعمل على احتمالية تشكيل التسلسل الرباعي للحروف فقط. وهذا خلاف نظام التشكيل بأدوات ماركوف الخفية إذ أضيفت له مجموعة من القوانين اللغوية للرفع من أداء التشكيل فيه.

العنصرين الجديدين في نظام التشكيل المستقل:

١. عنصر استخلاص الوحدات الرباعية لتسلسل الحروف من نص مشكل يدويا. هذه الطريقة تتسم بإختزال لا يؤثر على الكفاءة. بحيث أمكن الحصول على أقل من سبعين ألف تسلسل رباعي للحروف مع نسب احتمالية تشكيلها وهي كافية لاستخدامها لتشكيل نص عربي غير مشكل.

٢. عنصر اختيار تشكيل الحرف بناء على احتمالية وروده في أكثر من تسلسل رباعي مشكل. وهذا العنصر يتيح اختيار أفضل تشكيل ممكن للحرف عند تمرير التسلسلات الرباعية للحروف المشكلة واحتمالياتها على تسلسل حروف العبارة المراد تشكيل حروفها.

٤: الخلاصة:

قدم هذا التقرير تفاصيل مشروع المشكل الآلي والتقنيات التي أستخدمت في بناء ثلاثة نظم تشكيل مختلفة والنتائج التي وصلت إليها. ورغم ما توصل إليه الفريق من بناء نظام للتشكيل الآلي المستقل وما يتمتع به من سرعة ودقة في الأداء إلا أن المجال الآن مشرعا لمزيد من البحوث والتجارب لرفع كفاءة النظام خاصة وأن برمجياته معروفة ومتوفرة وتعتبر البنية الأولى للتشكيل، فلا تحتوي على مكونات أخرى كالقواعد الصرفية أو النحوية أو الدمج بين التسلسلات المختلفة. كل هذه وغيرها يمكن إضافتها لرفع نسبة التشكيل الصحيح. ونحن نقدم هذا التقرير لنأمل ألا نكون قد وضعنا أمام الباحثين تقنية جديدة للتشكيل الآلي للنص العربي ليس لتوظيفها والاستفادة منها وتطويرها فحسب ولكن نأمل أيضا أننا قد فتحنا آفاقا جديدة للتعامل مع النص العربي ومعالجته.

- Ananthakrishnan, Sankaranarayanan. Shrikanth S. Narayanan and Srinivas Bangalore (2005) Automatic Diacritization of Arabic Transcripts for Automatic Speech Recognition. In Proceedings of International Conference On Natural Language Processing. Kanpur. India.
- Chelba, Ciprian and Frederick Jelinek (2000) Structured language modeling. *Computer Speech and Language*. 14. 283–332.
- Dimitra, Vergyri and Katrin. Kirchho (2004) Automatic diacritization of Arabic for Acoustic Modeling in Speech Recognition. In Proceedings of the 20th International Conference on Computational Linguistics. Geneva. 66-73.
- Elshafei, Mustafa. Husni Almuhtasib and Mansour Alghamdi (2006). Machine Generation of Arabic Diacritical Marks. *The 2006 World Congress in Computer Science Computer Engineering. and Applied Computing* Las Vegas. USA. 26-29/6/2006.
- Elshafei, Mustafa. Husni Almuhtasib and Mansour Alghamdi (2006) Statistical Methods for Automatic Diacritization of Arabic Text. *The Saudi 18th National Computer Conference* Riyadh. 18: 301-306.
- Goweder, Abduelbaset. and Anne De Roeck Chelba. Ciprian and Frederick Jelinek (2001) Assessment of a Significant Arabic Corpus.
- Gal, Ya'akov (2002) An HMM Approach to Vowel Restoration in Arabic and Hebrew. *Computational Approaches to Semitic Languages*. University of Pennsylvania.
- Kontorovich, Leonid and Daniel D. Lee (2001) Learning Semitic languages with Hidden Markov Models. NIPS 2001 Workshop on Machine Learning Methods for Text and Images.
- Mustafa, Sulieman Hussein (1998) Arabic string searching in the context of character code standards and orthographic variations. *Computer Standards and Interfaces*. 20: 31-51.
- Nelken. Rani and Stuart M. Shieber (2005) Arabic Diacritization Using Weighted Finite-State Transducers. In Proceedings of the Workshop on *Computational Approaches to Semitic Languages Workshop*. University of Michigan. Ann Arbor. 79-86.
- Smrž, Otakar and Petr Zemánek (2002) Shards from an Arabic Treebanking Mosaic.

الملاحق:

الملحق ١:

يوجد في اللغة العربية ١٤ علامة تشكيل، تظهر هذه العلامات في ثلاثة مواضع في الكلمة العربية هي البداية والوسط والنهاية. الجدول التالي يوضح رموز علامات التشكيل التي قمنا باستخدامها في نظام التشكيل باستخدام أدوات ماركوف الخفية.

الرقم	علامة التشكيل	الموديل
١	الفتحة	mfa
٢	الضمة	mda
٣	الكسرة	mka
٤	الشدّة + فتحة	msf
٥	الشدّة + ضمة	msd
٦	الشدّة + كسرة	msk
٧	تنوين فتح	mtf
٨	تنوين كسر	mtk
٩	تنوين ضم	mtd
١٠	سكون	mso
١١	شدّة + تنوين فتح	stf
١٢	شدّة + تنوين ضم	std
١٣	شدّة + تنوين كسر	stk
١٤	بدون تشكيل	non

الملحق ٢:

ملاحظات على الجدول التالي:

١. العلامة ⊗ تعني: تسمح العربية بورد الحرف في هذا المكان من الكلمة
٢. العلامة ⊘ تعني: يمكن أن يرد الحرف بدون تشكيل
٣. الحرف "ب" في العمود تحت "الموضع" تعني في بداية الكلمة.
٤. الحرف "و" في العمود تحت "الموضع" تعني في وسط الكلمة.
٥. الحرف "ن" في العمود تحت "الموضع" تعني في نهاية الكلمة.
٦. "الرمز الخاص" في نهاية الجدول هو إشارة الصمت silence حيث هذا أحد متطلبات التعامل مع الإشارة الصوتية للكلام التي تبدأ بصمت وتنتهي بصمت.

علامات التشكيل													⊘	⊗	رقم الحرف	الموضع	الحرف	
std	stk	stf	msd	msk	msf	mso	mtd	mtk	mtf	mda	mka	mfa						
															x	١	ب	ء
x	x	x	x	x	x	x	x	x	x	x	x	✓	x	✓	٢	و	ء	
x	x	x	x	x	x	x	✓	✓	✓	✓	✓	✓	x	✓	٣	ن	ء	
														x	٤	ب	ؤ	
x	x	x	✓	x	✓	✓	x	x	✓	✓	x	✓	x	✓	٥	و	ؤ	
x	x	x	x	x	x	✓	✓	✓	x	✓	x	✓	x	✓	٦	ن	ؤ	
														x	٧	ب	ئ	
x	x	x	x	✓	x	✓	x	x	✓	✓	✓	✓	x	✓	٨	و	ئ	
x	x	x	x	x	x	✓	✓	✓	x	✓	✓	✓	x	✓	٩	ن	ئ	
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	✓	✓	١٠	ب	ا	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١١	و	ا	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١٢	ن	ا	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١٣	ب	آ	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١٤	و	آ	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١٥	ن	آ	
x	x	x	x	x	x	x	x	x	x	✓	x	✓	x	✓	١٦	ب	أ	
x	x	x	x	x	✓	✓	x	x	✓	x	x	✓	x	✓	١٧	و	أ	
x	x	x	x	x	x	✓	✓	✓	x	✓	✓	✓	x	✓	١٨	ن	أ	
x	x	x	x	x	x	x	x	x	x	x	✓	x	x	✓	١٩	ب	إ	

علامات التشكيل													∅	⊗	رقم الحرف	الموضع	الحرف
std	stk	stf	msd	msk	msf	mso	mtd	mtk	mtf	mda	mka	mfa					
x	x	x	x	x	x	x	x	x	x	x	✓	x	x	✓	٢٠	و	ا
x	x	x	x	x	x	x	x	✓	x	x	✓	x	x	✓	٢١	ن	ا
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٢٢	ب.ب	ب
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٢٣	و	ب.ب
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٢٤	ن	ب.ب
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٢٥	ب.ب	ت
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٢٦	و	ت
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٢٧	ن	ت
														x	٢٨	ب.ب	ة
														x	٢٩	و	ة
x	x	x	x	x	x	x	✓	✓	✓	✓	✓	✓	x	✓	٣٠	ن	ة
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٣١	ب.ب	ث
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٣٢	و	ث
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٣٣	ن	ث
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٣٤	ب.ب	ج
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٣٥	و	ج
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٣٦	ن	ج
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٣٧	ب.ب	ح
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٣٨	و	ح
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٣٩	ن	ح
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٤٠	ب.ب	خ
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٤١	و	خ
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٤٢	ن	خ
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٤٣	ب.ب	د
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٤٤	و	د
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٤٥	ن	د
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٤٦	ب.ب	ذ
x	x	x	x	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٤٧	و	ذ
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٤٨	ن	ذ

علامات التشكيل													∅	⊗	رقم الحرف	الموضع	الحرف
std	stk	stf	msd	msh	msf	mso	mtd	mtk	mtf	mda	mka	mfa					
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٤٩	ب.	ر
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٥٠	و	ر
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٥١	ن	ر
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٥٢	ب.	ز
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٥٣	و	ز
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٥٤	ن	ز
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٥٥	ب.	س
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٥٦	و	س
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٥٧	ن	س
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٥٨	ب.	ش
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٥٩	و	ش
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٦٠	ن	ش
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٦١	ب.	ص
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٦٢	و	ص
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٦٣	ن	ص
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٦٤	ب.	ض
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٦٥	و	ض
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٦٦	ن	ض
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٦٧	ب.	ط
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٦٨	و	ط
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٦٩	ن	ط
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٧٠	ب.	ظ
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٧١	و	ظ
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٧٢	ن	ظ
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٧٣	ب.	ع
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٧٤	و	ع
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٧٥	ن	ع
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٧٦	ب.	غ
x	x	x	x	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٧٧	و	غ

علامات التشكيل													∅	⊗	رقم الحرف	الموضع	الحرف
std	stk	stf	msd	msk	msf	mso	mtd	mtk	mtf	mda	mka	mfa					
x	x	x	x	x	x	✓	✓	✓	x	✓	✓	✓	x	✓	٧٨	ن	غ
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٧٩	ب.ب	فا
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٨٠	و	فا
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٨١	ن	فا
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٨٢	ب.ب	قا
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٨٣	و	قا
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٨٤	ن	قا
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٨٥	ب.ب	كا
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٨٦	و	كا
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٨٧	ن	كا
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٨٨	ب.ب	س
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٨٩	و	س
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٩٠	ن	س
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٩١	ب.ب	م
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٩٢	و	م
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٩٣	ن	م
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٩٤	ب.ب	ن
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٩٥	و	ن
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٩٦	ن	ن
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	٩٧	ب.ب	هـ
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	x	✓	٩٨	و	هـ
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	x	✓	٩٩	ن	هـ
x	x	x	x	x	x	x	x	x	x	✓	✓	✓	x	✓	١٠٠	ب.ب	و
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	✓	✓	١٠١	و	و
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	✓	✓	✓	✓	✓	١٠٢	ن	و
														x	١٠٣	ب.ب	ي
														x	١٠٤	و	ي
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	✓	١٠٥	ن	ي
x	x	x	x	x	x	x	x	x	x	✓	x	✓	x	✓	١٠٦	ب.ب	ي

علامات التشكيل													∅	⊗	رقم الحرف	الموضع	الحرف
std	stk	stf	msd	msk	msf	mso	mtd	mtk	mtf	mda	mka	mfa					
x	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	✓	✓	107	و	ي
✓	✓	✓	✓	✓	✓	✓	✓	✓	x	x	x	✓	✓	✓	108	ن	ي
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	x	109	فراغ	
x	x	x	x	x	x	x	x	x	x	x	x	x	✓	x	110	رمز خاص	

الملحق: ٣

١١٣٢٨	.	١	!	٢٣١٤٧٠	.	.	#
٢	.	.	!	١٩٨٩	.	.	ء
٥٩	.	١	آ	١٦٨	.	.	ء
١٢٠٨	.	.	آ	٢٠٤	.	.	ء
١٦٧٩٧٠	.	.	ا	٢٧٠	.	.	ء
١	.	.	ا	٥٣٥	.	.	ء
١	.	.	ا	٩٢٢	.	.	ء
١٧٥٣	.	.	ا	٢٧	.	.	ء
١٨٨	.	.	ا	٦	.	.	ء
٧	.	.	ا	٢	.	.	ق
١	.	.	ب	٣٧٥	.	.	ق
١٢١	.	.	ب	٧٣٢	.	.	ق
٣٩٨	.	.	ب	١١	.	.	ق
٤	.	.	ب	١٥٩	.	.	ق
٤	.	.	ب	١٢	.	.	ئ
٤٩٩	.	.	ب	١٧	.	.	ئ
١٢٨٨٩	.	.	ب	٣٢٨٩	.	.	ئ
١٥٥٨٠	.	.	ب	٥٦	.	.	ئ
١٦٣	.	.	ب	٦٨	.	.	ئ
٢٧٦١	.	.	ب	٧٧	.	.	ئ
٣٠١٥	.	.	ب	٩٤١	.	.	ئ
٣٠٥	.	.	ب	١٠١٥	.	.	أ
٣٢٣	.	.	ب	١٤	.	.	أ
٥١	.	.	ب	٢١	.	.	أ
١٩٧٥٣	.	.	ة	٢١٥٠	.	.	أ
١٥٦٧	.	.	ة	٢١٨٩٥	.	.	أ
٢٦٧٧	.	.	ة	٣	.	.	أ
٣٠١١	.	.	ة	٧	.	.	أ
٤	.	.	ة	٤٣	.	.	أ
٤٦٢٨	.	.	ة	٧	.	.	!

٢٠٣٥	•	•	ج	٤٧٢٢	•	•	٥٥
٢١	•	•	ج	٢٢	•	•	٥٥
٢٨٧٥	•	•	ج	١	•	•	ج
٣٧٠٧	•	•	ج	٢١٧	•	•	ج
٤٩	•	•	ج	٥٦٠١	•	•	ج
٦٨١٤	•	•	ج	٦	•	•	ج
٨١	•	•	ج	٧٤٥	•	•	ج
٤٣	•	•	ج	١٢٤٤	•	•	ج
١	•	•	ح	١٦٥٩	•	•	ج
٢١٨	•	•	ح	٢١٦٩٠	•	•	ج
٢٧٨	•	•	ح	٢٥٥	•	•	ج
٢٩	•	•	ح	٢٦٣٠	•	•	ج
١٣٧	•	•	ح	٤٥	•	•	ج
١٨٨٣	•	•	ح	٢٧٨٢	•	•	ج
٣٣٨٦	•	•	ح	١٠	•	•	ج
٤٢١٣	•	•	ح	٢٥٤	•	•	ج
٥٧	•	•	ح	٦٨	•	•	ج
٩٣	•	•	ح	٨١١	•	•	ج
٩٣١٥	•	•	ح	١٠٥	•	•	ج
٦	•	•	ح	١٤١١	•	•	ج
١	•	•	خ	١٦٣٨	•	•	ج
١	•	•	خ	٢٢	•	•	ج
٤٦	•	•	خ	٢٥٤٨	•	•	ج
٤٦	•	•	خ	٥٥	•	•	ج
٥٥	•	•	خ	١٠٩	•	•	ج
١٧٣٨	•	•	خ	٧	•	•	ج
١٩	•	•	خ	١٤٥	•	•	ج
٣١٧٨	•	•	خ	٢	•	•	ج
٣٨٧١	•	•	خ	٤٧	•	•	ج
٥	•	•	خ	٥	•	•	ج
٦	•	•	خ	٩١	•	•	ج

٦٢٠	•	•	ر	٩١٤	•	•	خ
١٢٣٧٣	•	•	ر	١٠١	•	•	د
١٧٤٢١	•	•	ر	١١	•	•	د
٣٨٥	•	•	ر	١٦٥٤	•	•	د
٥٣٠٩	•	•	ر	٢٣٠٢	•	•	د
٥٥٥١	•	•	ر	٥١	•	•	د
٧١٧	•	•	ر	٨٧٥	•	•	د
٨٧٢	•	•	ر	١١٠٢١	•	•	د
١٠	•	•	ر	٢٣٢	•	•	د
٢	•	•	ز	٣١٩٩	•	•	د
٣٤٧	•	•	ز	٣٥٠٩	•	•	د
٤١١	•	•	ز	٤٠٢	•	•	د
٨٧	•	•	ز	٦٢٦	•	•	د
١٦٧٢	•	•	ز	٧٤٣٣	•	•	د
١٧٠٨	•	•	ز	٣٩	•	•	د
١٨	•	•	ز	١٣٠	•	•	ذ
٤٥	•	•	ز	٢	•	•	ذ
٥٣	•	•	ز	٣٨٦	•	•	ذ
٥٤٣	•	•	ز	٨٤	•	•	ذ
٧٢٨	•	•	ز	٢٨	•	•	ذ
٥	•	•	ز	٣	•	•	ذ
١٣٩٦	•	•	س	٣٥٠٩	•	•	ذ
٢	•	•	س	٤٤	•	•	ذ
٢	•	•	س	٤٥٨	•	•	ذ
٧٠٨	•	•	س	٥٣٩	•	•	ذ
٨٠٠	•	•	س	٥٤٨٥	•	•	ذ
١١٤	•	•	س	١٠٧٣	•	•	ر
١٦١٨	•	•	س	١٦	•	•	ر
٤٤٦٣	•	•	س	١٩٢٤	•	•	ر
٥١	•	•	س	٢٤	•	•	ر
٧٢١٢	•	•	س	٢٧	•	•	ر

۱۰۰	ض	۷۵۸۹	ص
۴۰	ض	۹۶	ص
۴۰۷	ض	۷	ص
۱۵۰	ض	۱	ص
۲۶۱۵	ض	۱	ص
۲۶۴۷	ض	۱۳۵۴	ص
۳۶۳	ض	۱۴۹	ص
۴۸	ض	۲	ص
۷۲۱	ض	۲۵۸	ص
۹۶۹	ض	۲۰۰	ص
۱	ط	۲۳	ص
۱۰۹۹	ط	۳	ص
۱۲	ط	۳۶۹۲	ص
۲۲۸	ط	۴۵۲	ص
۴۱۸	ط	۹	ص
۱۰۹	ط	۹۳۴	ص
۱۴۹۹	ط	۱	ص
۲۲۶۳	ط	۱۱	ص
۳۹۱۲	ط	۱۱۵۲	ص
۵۵	ط	۱۴	ص
۵۹	ط	۳۶	ص
۸۵۹	ط	۳۸۲	ص
۳	ط	۸۸۳	ص
۲	ظ	۱۲۸۵	ص
۲۳۴	ظ	۱۸۹۲	ص
۵۴	ظ	۲۰	ص
۸۶	ظ	۲۱۲۹	ص
۱۰	ظ	۳۰۷۰	ص
۱۵۲۹	ظ	۶۷	ص
۱۶	ظ	۷۳	ص
۲۵۴	ظ	۲	ص

١٠١	.	*	فا	٥٠٦	.	ظ
١٠٦١٧	.	'	فا	٥٢٨	.	ظ
١٣٣٩٣	.	'	فا	٧	.	ظ
١٦٩	.	"	فا	١	.	ظ
١٩٢٤	.	'	فا	١	"	ع
٢٤٧٠	.	°	فا	١٢	"	ع
٣١٤	.	"	فا	١٦٩	'	ع
٥	.	°	فا	٧	'	ع
١١٤	'	'	فا	١٣٥	.	ع
١٣	'	'	فا	٢٠٣٦٠	.	ع
١٤	"	'	فا	٣١٥٨	.	ع
٢٠٢	"	'	فا	٣٣٦	.	ع
٣٩١	"	'	فا	٣٤٧	.	ع
٨	"	'	فا	٦٥٢٦	.	ع
١٠٤٦٤	.	'	فا	٦٨٦٩	.	ع
١٧٤	.	"	فا	٢	.	ع
٢١٤	.	"	فا	٤	"	غ
٢٧٠٣	.	'	فا	٨	"	غ
٣٢٣٢	.	°	فا	١١٣٢	.	غ
٥٠٢٠	.	'	فا	٢٦٢١	.	غ
٧٣	.	'	فا	٣٤٣	.	غ
٢	.	°	فا	٥	.	غ
١	"	'	ك	٧	.	غ
١١	"	'	ك	٧٤٨	.	غ
١٣٠	'	'	ك	٩	.	غ
٢	'	'	ك	١	"	فا
٢٦١	"	'	ك	١	'	فا
٥٧٣	"	'	ك	١١	"	فا
١١٤٢١	.	'	ك	١٣٦	"	فا
٢٦٥١	.	°	ك	٢٤٣	'	فا
٣٣	.	'	ك	٨٤	'	فا

٦٠٤٨	٠	٠	م	٣٤	٠	ك
٤٣	٠	٠	م	٣٦٣٦	٠	ك
٢٣	٠	٠	ن	٣٧٥٨	٠	ك
٣	٠	٠	ن	٥٥	٠	ك
٤٥٧	٠	٠	ن	٥	٠	ك
٦	٠	٠	ن	١٢١٢	٠	ل
٦٨٥٠	٠	٠	ن	٢٤	٠	ل
٨٤٨	٠	٠	ن	٤٤	٠	ل
١٤١	٠	٠	ن	٥٨٤٢	٠	ل
١٨١٥١	٠	٠	ن	٧٨٤	٠	ل
١١٩٩١١	٠	٠	ن	٨٠	٠	ل
٢٨٠	٠	٠	ن	٢٢٥٥٩	٠	ل
٣٧٠٠	٠	٠	ن	٢٦١٢٢	٠	ل
٤٨٠	٠	٠	ن	٢٦٥	٠	ل
٩٠٧٠	٠	٠	ن	٤٨٦٤٤	٠	ل
٢٤	٠	٠	ن	٥١٣١	٠	ل
١٩	٠	٠	ه	٧٠١	٠	ل
٣٣	٠	٠	ه	٧٠٥	٠	ل
٤٦	٠	٠	ه	٢٢٢٢١	٠	ل
١٣٦١٦	٠	٠	ه	١٨	٠	م
١٥٧١	٠	٠	ه	٢٥٥٢	٠	م
٢٧	٠	٠	ه	٤٧٠	٠	م
٣٢	٠	٠	ه	٥٥	٠	م
٥٥	٠	٠	ه	٥٩	٠	م
٧٩٧٢	٠	٠	ه	٦٢٤	٠	م
٩٢٤٨	٠	٠	ه	١٤٩٩٢	٠	م
٢٦	٠	٠	ه	١٨٢٢٣	٠	م
١	٠	٠	و	١٩٢	٠	م
١٠٦٨	٠	٠	و	٢٥٦٢٤	٠	م
١١	٠	٠	و	٣٩٣	٠	م
١٧	٠	٠	و	٤٦٤	٠	م

٣٦١٠	'	'	ي	٣٢٦	'	'	و
٥١٣	'	'	ي	٤٧٦	'	'	و
٥٥٤	'	'	ي	٣١٧٨	'	'	و
١٥	'	'	ي	٣٢	'	'	و
١٥٣٣٢	'	'	ي	٣٤٦٠٤	'	'	و
١٧٨	'	'	ي	٤	'	'	و
٢٠	'	'	ي	٤٢	'	'	و
٤١٧٤	'	'	ي	٥٨٢٦	'	'	و
٧٣٣٢	'	'	ي	٨١٧	'	'	و
٧٤٣	'	'	ي	١٥٦٣١	'	'	و
١٠٠٦٤	'	'	ي	١١٢٩٠	'	'	ي
٣٥٤١٥	'	'	ي	١١٦٧	'	'	ي
				٢٥٤	'	'	ي