المحللات الصرفية للغة العربية

تقرير مقدم من د/ سلوى السيد حماده
للمنظمة العربية للعلوم والثقافة
hesalwa@yahoo.com

ملخص

بعد دخول الحاسوب في كل المجالات وكل البلدان،وتقدم برمجيات الحوسبة اللغوية للغة الإنجليزية خاصة ولغات أخرى مثل الألمانية والفرنسية واليابانية وغيرها. برزت الحاجة لتطويعه للّغات عالمية أخرى، ومنها اللغة العربية، وظهرت صعوبات في إمكانية تنفيذ ذلك على أنظمة لم تصمم أصلًا للغات مختلفة ولكنها تيسرت نسبيا بإضافة برمجيات وتقنيات ملحقة خاصة بثقافات ولغات أخرى، كانت اللغة العربية واحدة منها. ولا ننكر محاولات شركات القطاع الخاص بالمنطقة العربية في عمل هذه المحللات بمهارة حيث يتوفر لها الدعم بكافة صوره إلا أن الجهود البحثية مازالت تحبو ببطء لحاجة هذه الأعمال للدعم المادي والمعنوي والبشري.

(١)المقدمة

مر أكثر عشرون عامًا تقريبًا على ابتكار الحاسوب الشخصي، وعلى الرغم من أنه كان حصيلة قرون من التطور المعرفي والتقني الذي شاركت فيه الإنسانية جمعاء، إلا أن مبتكريه في الولايات المتحدة الأمريكية لم يأخذوا بالاعتبار أثناء تصميمه سوى حاجات أسواقهم وثقافتهم ولغتهم، ولم يأخذوا في اعتباررهم باقي لغات العالم. ولكن هذا الانتشار قد حدث بالفعل. ونحن في المنطقة العربية لم يعد أحد منا يقلق من جهة إمكانية التحرير والطباعة باللغة العربية، واستخدام واجهات برامج عربية، بل أصبح بإمكاننا استخدام قواعد بيانات، وجداول ممتدة، وصفحات ويب وبريد إلكتروني باللغة العربية، ومجموعة واسعة من البرامج التعليمية والقواميس، وبرمجيات الأعمال، لكن كل تلك الإمكانيات –على الرغم من أهميتها– ليست أكثر من كونها تعريبًا مباشرًا للبرامج الإنجليزية، أو تطبيقًا لتقنيات عالمية غير متعلقة مباشرةً باللغة العربية، ولا تدخل في عمق تلك اللغة، أو توفر الأدوات اللازمة للاستفادة الكاملة من ثرائها. إلا أننا بدأنا نشهد اعتبارًا من منتصف عقد التسعينات نتائج البحث والتطوير لتقنيات حاسوبية تهدف إلى معالجة اللغة العربية ذاتها، وتقدم حلولًا في مجالات الترجمة الآلية، والتدقيق الإملائي والنحوي، والتعرّف على الكلام وتحويله إلى نصوص، والتعرف الضوئي على النصوص، والتحليل الصرفي والبحث واسترجاع النصوص، وتوليد الخطوط العربية آليًا. لكن تلك التقنيات ما زالت في بداية الطريق، ولا تُستخدم على نطاق واسع؛ لأنها لا توفر حلولًا عملية تناسب الاستخدام على نطاق

واســـع في مجال الأعمال، كما لا تســتطيع تقديم ثقافة العالم للمســتخدمين العرب، وتقدم اللغة والثقافة العربية إلى العالم. وهذا لأن مطوري تلك التقنيات يواجهون معوقات كثيرة، من أهمها طبيعة اللغة العربية نظرا لثرائها واتساعها، كما تشكل قلة الأبحاث الأكاديمية التقنية المتعلقة بها معوقا ليس بالهين، مما يوجب على الشركات المطورة لتقنيات اللغة العربية إجراء أبحاث أكاديمية مكلفة في علوم اللغويات والرياضيات والصرف، ولهذه الأبحاث تكاليف باهظة لا يستطيع تحملها إلا الشركات الكبيرة، ومراكز البحث العلمي والجامعات.

وما يزيد الأمر صعوبةً أن اللغة العربية غنية جدًا بالمفردات والقواعد، يتضح هذا بمقارنة بسيطة لعدد المفردات بين اللغتين العربية والإنجليزية –مثلًا– فإذا علمنا أن عدد المفردات في اللغة الإنجليزية لا يتعدى الـــــ ٣٠٠ ألف كلمة، نجد أن عدد مفردات اللغة العربية يزيد على ســبعة ملايين كلمة.

(٢)المحلل الصرفي

هو برنامج يقوم بدراسة بناء الكلمة، وتحويلها إلى جذرها الأصلي.

(٣)طبيعة اللغة العربية:

من المعروف أن الأبحاث المتعلقة باللغة العربية يشوبه نقص خطير، وسبب هذا النقص الرئيسي أن اللغة العربية لغة صرفية لا لغة لصقية بمعنى أن أصل الكلمة فيها يتغير لعدة أشكال تختلف عن الشــكل الأصــلي للكلمة عكس اللغات اللاتينية فهي لغات لصــقية حيث يظل أصــل الكلمة «الجذر» ثابت ويتغير المعنى تبعًا لزيادة أو نقصــان حروف على هذا الأصــل «الجذر»، عكس العربية التي يحتاج البحث فيها للوصــول إلى نتائج دقيقة إلى تحليل صــرفي، فمثلًا في العربية: إذا بحثنا عن كلمة «عمل» فلن يظهر في نتائج البحث ســوى الكلمات التي تحتوي على أصــل هذه الحروف، مع احتمال ضياع نتائج هامة تشترك في نفس الجذر، ولكن نظرًا لاختلاف بنيتها يســـتحيل الوصـــول إليها بطرق البحث المطابقة العادية فمثلًا: لن نصـــل إلى كلمات مثل: «الأعمال، عاملون، عاملات» مع كونها من نفس أصل الكلمة المبحوث عنها بعكس المستخدم مثلًا: في الإنجليزية إذا بحث المســتخدم عن كلمة work فســوف تظهر له تصــريفات وأشــكال أخرى لهذه الكلمة نحو works أو worked وهكذا.

الهيكل البنائي للكلمات في اللغة العربية[١]

- اللغة العربية لغة سامية تبنى كلماتها من الجذور
- الكلمات العربية تمتاز بكثرة مشتقاتها
- النظام النحوي للغة العربية قائم على أســاس الجذور والأوزان الصــرفية للكلمات والتي لا تزيد عن ١٠٠٠٠ جذر و٩٠٠ وزن صرفي.

- تشـــكل الجذور الثلاثية الأغلبية العظمى لكلمات اللغة (٨٥%)، بينما تشـــكل الجذور الرباعية والخماسية والسداسية الأقلية من الكلمات حيث أن الأسماء والأفعال تتشكل من هذه الجذور عن طريق إضافة الزوائد الاشتقاقية إليها.

- تحدد التغيرات التي يبديها الجذر والأوزان الصـــرفية المعنى الفعلي للكلمة ("كتاب" على وزن فعال و "كاتب" على وزن فاعل).

- إن الكلمة في اللغة العربية قد تكون مؤلفة من الأصـــل (stem) وهو عبارة عن جذر صـحيح أو وزن صـرفي مضـافاً إليه بعض الزوائد. والزوائد هي الحروف المكونة للهيكل الصـرفي للكلمة بخلاف حروف الجذر، وتشمل العلامات الصـرفية للزمن والعدد وحالات التذكير والتأنيث. كما تتضـمن أيضـاً بعضـاً من حروف الجر والربط والمحددات لعمل الأسماء وضمائر الملكية والضمائر المتصلة والمنفصلة.

أمثلة[٢]:

- الضمائر والأحرف تتصل بالكلمة مباشرة

مثال "كتابه" "لكتابها" "والبيت" ... الخ

- لتحليل الصـور المتعددة للكلمة الواحدة يجب ازالة الأحرف الزائدة (اللواحق) من الكلمات وارجاعها إلى اصلها.

مثال : "كتابهم" "كتابها" "كتابه" "وكتابهما" وجذعهم باستخدام وجذرهم"كتب".

(٤)أهمية التحليل والمحلل الصرفي

تعتبر تقنيات التحليل الصـرفي Morphological Processor بمثابة اللب للتقنيات والتطبيقات العربية الأخرى؛ حيث تسـتفيد منها بشـكل أو بآخر، لكنها تصـير أسـاسـية جدًا بالنسـبة لتقنيات وإمكانيات البحث والفهرســة؛ بحيث يمكن إعتبارها تطبيقًا مباشــرًا لها. وتشــمل تقنيات التحليل الصـرفي إعادة الكلمة المشـتقة إلى جذرها، أو اشـتقاق جميع الكلمات الممكنة من جذر معين، أو تحليل الكلمة إلى عناصـرها الأولية، ويفصـل بين جذور الكلمات وسوابقها ولواحقها، ويجردها من صـورها الصـرفية، وهو أسـاس للعديد من المنتجات والتطبيقات والحلول التي تتطلب معالجة ذكية للغة العربية، كمحرك البحث في الإنترنت. ومن تطبيقاتة:

وهو على خلاف المحلل الصـــرفي الإنجليزي الذي يقف عند حدود ســـاق الكلمة/ جذع Stem، فإنه يستطيع استخلاص الجذر، ويتعرف على الصـورة الصـرفية للكلمة، فهو مثلًا، يفيد أنه يمكن اشـــتقاق المصـــدر «إيجاد»، من الفعل الثلاثي «وجد»، لأنه على وزن «إفعال». يمكن للمحلل

الصرفي أيضًا أن يولد كافة المشتقات الصحيحة للكلمة، كأن يولد الكلمات «آت ومؤتي ويأتون ومؤات وإتيان ويأتي ويأتين» من الفعل «أتى». وهو يفيد في استخلاص الكلمات المحورية من النصوص عند البحث فيها، لأنه يتعرف على لواصق الكلمة، فيكفي البحث عن «أكل»، مثلًا، لتأتي جميع مشتقاتها وصورها وأشكالها ضمن نتائج البحث.

يستخدم المحلل الصرفي في أغلب التطبيقات اللغوية خاصة:

- التحليل الصرفي للكلمات
- التوليد الصرفي اللكلمات
- الترجمة الآلية
- استرجاع البيانات

ومن هنا تتضح أهمية محلل صرفي كامل لكلمات اللغة العربية بحيث يتمكن المستخدم من الوصول إلى كافة أشكال الكلمة وتصريفاتها المختلفة إلى نتائج دقيقة لأي بحث مطلوب.

من المعروف أن معظم كلمات لغات العالم لها العديد من المعاني حيثُ يمكن استخدام الكلمة للتعبير عن أكثر من معنى ويحدد هذا المعاني موقع الكلمة داخل الجملة أو السياق، وتعد هذه المشكلة من المشكلات المعقدة أثناء المعالجة الآلية لتلك اللغات، فعملية فك لبس المعنى والمعروف باللبس الدلالي من أكبر الصعوبات والتحديات التي تواجه المتصدر لمعالجة لغة ما، هذا بشكل عام أما إذا تحدثنا عن اللغة العربية فسنجد الأمر أكثر صعوبة لما لهذه اللغة من طبيعة خاصة قد يصعب في بعض الحالات التي معها بطريقة سهلة لفك اللبس، فهناك المشاكل المتعلقة بتطبيقات الترجمة، واللبس الصرفي، ولبس المعنى، ومشكلة الضمائر التي تعود على عاقل أو غير العاقل، بعكس الإنجليزية التي تفرق بين العاقل وغير العاقل في الضمائر سواء كانت ضمائر سواء كانت ضمائر الفاعل أم المفعول أم الملكية. والتحليل الصرفي يساهم في حل هذه المشكلات أحيانا واحيانا أخرى يحتاج أن تحل حتى يكتمل.

أنواع الكلمات

تنقسم كلمات اللغة إلى عدة فئات:

- الفئة الأولى: جذر بدون إضافات
- الفئة الثانية: جذر يحتوي على لواحق فقط
- الفئة الثالثة: جذر يحتوي على بوادئ فقط
- الفئة الرابعة: جذر يحتوي على بوادئ ولواحق
- الفئة الخامسة: تصريف (مع إضافة وسطية) بدون بوادئ ولواحق
- الفئة السادسة: تصريف (مع إضافة وسطية) يحتوي على لواحق فقط
- الفئة السابعة: تصريف (مع إضافة وسطية) يحتوي على بوادئ فقط

الفئة الثامنة: تصريف (مع إضافة وسطية) يحتوي على بوادئ ولواحق

(٥)طرق التحليل الصرفي:

المحلل الصـــرفي العربي على خلاف المحلل الصـــرفي الإنجليزي الذي يقف عند حدود ســـاق الكلمة Stem/ جذع، فإنه يستطيع استخلاص الجذر، ويتعرف على الصورة الصرفية للكلمة، فهو مثلاً، يفيد أنه يمكن اشـــتقاق المصـــدر "إيجاد"، من الفعل الثلاثي "وجد"، لأنه على وزن "إفعال". يمكن للمحلل الصرفي أيضاً أن يولد كافة المشتقات الصحيحة للكلمة، كأن يولد الكلمات آت ومؤتي ويأتون ومؤات وإتيان ويأتي ويأتين من الفعل أتى. وهو يفيد في اســتخلاص الكلمات المحورية من النصـــوص عند البحث فيها، لأنه يتعرف على لواصـــق الكلمة، فيكفي البحث عن "كتب"، مثلاً، لتأتي جميع مشتقاتها وصورها وأشكالها ضمن نتائج البحث.

تعتبر تقنيات التحليل الصـــرفي مركز أغلب التقنيات والتطبيقات اللغوية ، التي تعتمد عليها بطريقة ما. وتعتبر أساسية بالنسبة للبحث في اللغة والفهرسة الآلية لأن كلاهما يعتمد عليها كليا بالنســبة للغة العربية. والتحليل الصــرفي يمكن من الوصــول لجذر الكلمة وتحديد المشــتقات الصــرفية. هناك عدد من الأبحاث التي تناولت حل مشــاكل الصــرف العربي خاصــة في مجال التحليل الصــرفي، وذلك لأغراض البحث في النصــوص واســترجاع المعلومات. وقد اتبعت هذه الدراســـات طرقا مختلفة لمعالجة الكلمات صـــرفيًا. وبُنيت بعض هذه الأبحاث على المدارس الحديثة في علم الصرف Morphology وهذه الأبحاث مقسمة إلى أربع أقسام كالتالي[3]:

١. طريقة قوائم الكلمات المخزنة Table Lookup

تخزن جميع الكلمات في قوائم مع مكوناتها الصــرفية على شــكل جدول كبير. ويتم تحليل الكلمة بالبحث عنها في هذا الجدول ومن ثم معرفة جذرها ببســـاطة. ويمكن تطبيق هذه الطريقة على نصوص معينة مثل: القرآن الكريم محدد، أو مجموعة كتب محددة.

٢. الطريقة اللغوية Linguistics

تحويل قواعد اللغة العربية الصـــرفية المعروفة إلى خوارزميات حاســـوبية. وذلك بمحاكاة عمل اللغوي عند تصريف الكلمات واشتقاقها.

٣. الطريقة الرياضية Combinatorial

تحليل الكلمة بشـــكل آلي بطريقة التجربة والخطأ والتصـــحيح وذلك باعتبار الكلمة مجموعة من الحروف. حيثُ يتم أخذ ثلاثة أحرف منها ومقارنتها بقائمة من الجذور فإذا لم يوجد الجذر في القائمة أخذت ثلاثة أحرف أخرى... وهلم جرا إلى أن يتم إيجاد الجذر الأقرب إلى الصواب.

٤. طريقة الأوزان Pattern- Based

توليد مجموعة كبيرة من القواعد النصية الآلية Regular Expressions عن طريق المقارنة بين قائمة كبيرة من الكلمات مع ما يقابلها من مصادرها. ويتم استخدام هذه القواعد لمعرفة مصدر الكلمة، فإذا انطبقت على الكلمة أكثر من قاعدة واحدة يتم ترجيح أكثر القواعد تكرارًا."

(٦)المشذبات العربية [١]:

- التشذيب (Stemming) هو عملية إزالة كل من بوادئ الكلمة ولواحقها لإنتاج الجذر أو الأصل أو الجذع stem.
- التشذيب هو طريقة رياضية تجمع كل الكلمات التي تتقاسم نفس الأصل وتملك بعض العلاقات الدلالية.
- تصنف خوارزميات التشذيب العربية طبقاً لمستوى التحليل المطلوب: على أساس الجذع أو على أساس الجذر.
- طرق التشذيب المبنية على أساس الأصل تعمل على إزالة كل من البوادئ واللواحق الملحقة بالكلمة، بينما تحول الطرق المبنية على أساس الجذر كل من الأصول إلى جذور.
- إن المشذب العربي يشير إلى إزالة وتجريد الكلمة من الزوائد المضافة دون المحاولة للتعامل مع الإضافات الوسطية أو التعرف على الأوزان الصرفية واكتشاف الجذور للكلمات.

(٧)متطلبات عمل المحلل[٤]

١. قائمة بجميع السوابق الممكنة. (Prefix) ال التعريف، تاء الفاعل،... إلخ
٢. قائمة بجميع اللواحق الممكنة. (Suffix) واو الجماعة، نون النسوة،... إلخ
٣. قائمة بجميع الأوزان الأساسية. (Patterns) فعل، فعول، مفاعيل،... إلخ
٤. قاعدة بيانات بجميع الجذور الثلاثية منها و الرباعية.
٥. قاعدة بيانات بجميع الكلمات الجامدة.

تحليل الكلمة:

تحلل الكلمة باستخدام برمجيات عديدة منها لغة سي ++ (C++) للبرمجة. يقوم البرنامج بتحليل الكلمة وتحديد السوابق واللواحق والجذر.

(٨)تطبيق طرق التحليل الصرفي[٤]:

المشتقات و الجوامد:

– المشتق: و هو المولد أو المتفرع من جذر، مثل: مكتبة مأخوذه من الجذر (ك ت ب).

- الجامد: و هي التي ليس لها جذر، مثل: بئس، نِعم، إنسان ...

- **الجذر:** و طولة من ٣-٤ أحرف و الجذر هو الحروف الأصـــلية في الكلمة، و الجذر ليس بكلمة و إنما مجموعة من الحروف.
- **الوزن:** الوزن = الجذر + حروف أو حركات ضبط بالشكل مزيدة

- **السوابق:** الحروف الزائدة في بداية الكلمة.

- **اللواحق:** الحروف الزائدة في نهاية الكلمة.

الفرق **بين الجذع stem والجذر root** :

الجذع: هو المتبقي من الكلمة بعد حذف السوابق prefixes واللواحقsuffixes .

مثال

١. "**أفبالباطل**": هذه الكلمة لها أربع ســوابق هم (أ – ف – ب – ال) وليس لها أية لواحق وعليه يصح الجذع هو(باطل).

٢. " **فسيكفيكهم** ": هذه الكلمة لها سابقتين (ف – س) ولاحقتين هما(ك- هم) وعليه يصح الجذع هو(يكفي).

الجذر: هو أصـل الكلمة. أي الجزء المتبقي بعد حذف الأحرف الغير أصـلية من سـوابق ولواحق وأواسط. وللغة العربية أكثر من ٧٠٠٠ جذر تقريباً.

مثال:

١. "**أفبالباطل**": هذه الكلمة لها أربع ســوابق هم (أ – ف – ب – ال) وليس لها أية لواحق وحرف وسطي هو(ا). وبحذف هذه الأحرف يصبح الجذر(بطل).

٢. " **فســيكفيكهم** ": هذه الكلمة لها ســابقتين (ف – س) وحرف مضــارعة (ي) ولاحقتين هما(ك- هم)وبحذف هذه الأحرف يصبح الجذر(كفي).

وعلى هذا سنعتمد الطريقتين:

١) التجذير

أ. التجذير وإزالة الزوائد

تعتمد طرق التحليل الصـرفي على إزالة الزوائد من سـوابق ولواحق وأواسـط–حروف في وسـط الكلمة–، وازالة اللواحق من الكلمات العربية إذا تمت بصـورة عشـوائية يؤدي إلى الوصـول لجذر غير صحيح للكلمة. انظر الجدول التالي[2]:

الكلمة	اللاحقة المزالة	الجذر
وطنها	الواو والهاء	طن
وطن	لا شي	وطن
ومال	الوالو	مال
كمال	الكاف	مال
مالك	الكاف الملكية	مال
والتر	الواو واداة التعريف	تر

ب. أنواع انظمة إزالة الزوائد

- انظمة الازالة الخفيفة light stemmers
  - تقوم بإزالة اللواحق من أول وآخر الكلمة
  - ومدارسهم ---- مدارس

ومن أمثلته **نظام ليه لاركي Leah Larkey**

- يستبدل كلا من ''أ''،''إ''،''آ'' بـ ''ا''
- يستبدل ''ى'' بـ ''ي'' و ''ة'' بـ ''ه''
- يحذف حرف الواو من اول الكلمة على ان يبقى ثلاثة احرف او اكثر
- يحذف "الـ" ، "والـ" ، "فالـ" ، "كالـ" ، "بالـ"، "للـ" على ان يبقى حرفان او اكثر
- يحذف "ها"، "ان"، "ات"، "ون"، "ين"، "يه"، "ية"، "هـ"، "ة"، "ي" على الترتيب على ان يبقى حرفان او اكثر

مثال

- والدها
  ١. ازالة "و" يتبقي "الدها"
  ٢. ازالة "الـ" يتبقى "دها"

٣. ازالة "ها" لا يمكن "دها"

النتيجة هي "دها"

- ودها

١. ازالة الواو يتبقى "دها"

النتيجة هي "دها"

- انظمة الازالة الثقيلة heavy stemmers
  - تقوم هذه الانظمة بازالة اللواحق ثم ارجاعها الى الجذريمقارنتها بالاوزان الصرفية
  - ومدارسهم ---- مدارس ---- درس

ومن أمثلته **نظام شرين خوجة Shereen Khoja**[5]

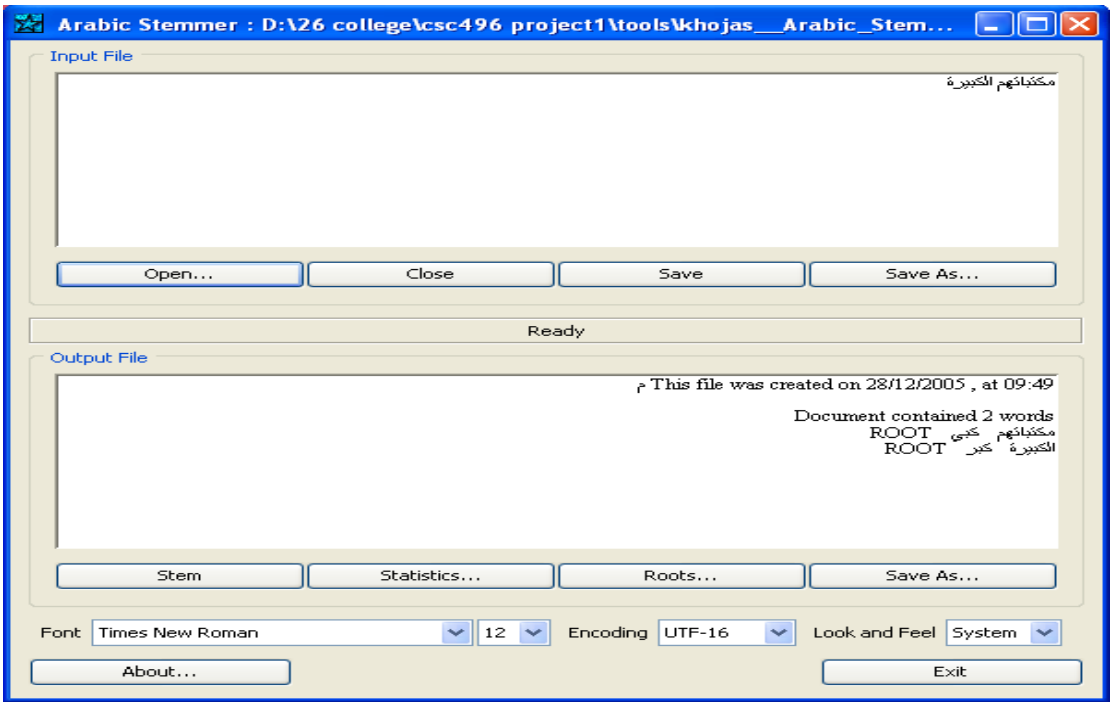
كتب هذا البرنامج بلغة بلغة جافا Java **وهو يأخذ الكلمة كمدخل ويخرج المخرج كما بالشـــكل التالي. وعيبه أنه يولد الجذر لا الجذع.**

كيفية العمل

- بحدف اداة التعريف "ال"،"وال"،"كال"،"فال"،"بال"، "لل".
- بحدف حرف "و" من اول الكلمة على ان يبقى ثلاثة احرف أو أكثر.
- بحدف حرف "ة" من أخر الكلمة.
- يحدف باقي احرف الجر والعطف من اول الكملة.
- يحدف باقي اللواحق التي تلحق اخر الكلمة.
- في كل مرة بعد حدف لاحقة يقارن النظام باقي الكلمة بقائمة الأوزان الصـرفية التي لها نفس عدد الاحرف.
- اذا طابقة الكلمة المتبقية ميزانا صـــرفيا يرجع الجدر ويختبر وجوده في قائمة الجذور الصحيحة.
- اذا لم يوجد الجذر ارجع الكلمة الأصلية والا يتم ارجاع الجذر.
- مثال والدها:
  - ازالة "و" يتبقي "الدها" لا يوجد ميزان صرفي.
  - ازالة "ال" يتبقى "دها" الميزان الوحيد "فعل".
  - "دها" لا توجد في قاموس الجذور.
  - النتيجة النهائية هي "والدها".
- والد

- يتم مقارنتها بالاوزان الصرفية لانها أربعة أحرف.
- الميزان الوحيد هو "فاعل" وذلك لمطابقة الألف الوسطى.
- النتيجة هي ما يناظر الاحرف "ف"، "ع"،"ل" في الكلمة.
- النتيجة النهائية هي "ولد" وهو موجود في قاموس الجذور.

- مكتباتهم الكبيرة انظر الشكل التالي للمخرج



شكل() مخرج البرنامج

ت. مزايا وعيوب هذه الانظمة[٢]:

- انظمة الازالة الخفيفة
  - ✓ سريعة وبسيطة لا تحتاج الى أي تحليل نحوي
  - ✓ تقع في اخطاء كثيرة فهي تجذر بصورة خاطئة
- انظمة الازالة الثقيلة
  - ✓ تقلل من حجم الفهرس بصورة كبيرة
  - ✓ أقل دقة من انظمة الازالة الخفيفة وذلك لعموم معني الجذر
  - ✓ تحتاج الى وقت اكبر وذلك للتحليل الكامل للكلمة
  - ✓ يصعب تحديد الجذر بدقة وذلك لغياب التشكيل

- اثبتت الابحاث الســابقة ان انظمة الازالة الخفيفة تعطى نتائج افضــل من انظمة الازالة الثقيلة.

ث. كيف يتم الازالة باستخدام انظمة الازالة الخفيفة[٢]:

- يتم ازالة اللواحق عن طريق مقارنة بداية ونهاية الكلمة بقائمة معدة مســـبقا مع شرط ان يكون الاحرف المتبقية اثنان او اكثر

والمغاربة	والـ	والمغاربة	هم	مغاربة	فعل	مغارب		غرب
	بالـ		هما		فاعل			
	الـ		مغاربة		مغارب		غرب	
	ل		ة		فعول			
			ية		مفاعل			
			كما		مفعول			
	ب		ي		فواعل			

(٩)خطوات عمـل المحللات الصـــرفيــة Morphological Analysers:

كا أوضحنا أن وظيفة المحلل الصرفي morphological analyzer هي تحديد جذور الكلمات والمعلومات الصـــرفية النحوية morpho-syntactic ذات العلاقة بهم. وللمحللات الصـــرفية أهمية بالغة في معالجة اللغة الطبيعية. ومشـكلة العربية غياب علامات التشـكيل والتي تعد جزءً أساسيا من أساسيات الكلام. وانظر إلى الجملة الإنجليزية الآتية:

Ths s wht nglsh lks lk wtht vwls.

هل تستطيع قراءتها في غياب حروف المد. ولكن انظر لها بعد استرجاعهم:

This is what English looks like without vowels.

وذلك هو حال العربية بعد رفع علامات التشكيل أي تزيد بها احتمالية اللبس بين الكلمات.

الخطوات[٤]

١. ابحث عن الكلمة في قائمة الكلمات الجامدة إذا وجدت: انتهى؛ إذا لم توجد: اذهب إلى الخطوة الثانية؛
٢. بإستخدام قائمة السوابق واللوحق احذف جميع السوابق واللوحق المتصلة بالكلمة؛
٣. اختر الوزن المناسب للكلمة
٤. استخرج الجذر من الكلمة

لنأخذ كلمة (المرابطون)

- **أولاً:** نبحث عن الكلمة في قائمة الكلمات الجامدة

المدخلات	ا	ل	م	ر	ا	ب	ط	و	ن
الوزن									
الجذر									

ثانياً: نحذف جميع السوابق واللوحق المتصلة بالكلمة:

المدخلات	ا	ل	م	ر	ا	ب	ط	و	ن
الوزن									
الجذر									

ثالثاً: من قائمة الأوزان نختار الوزن المناسب للكلمة:

- نفس عدد الأحرف.
- تطابق الأحرف ما عدا – ف ع ل –

المدخلات	ا	ل	م	ر	ا	ب	ط	و	ن
الوزن			م	ف	ا	ع	ل		

الجذر									

رابعاً: نستخرج الجذر من الكلمة:

- باختيار الأحرف المقابلة لـ – ف ع ل –

المدخلات	ا	ل	م	ر	ا	ب	ط	و	ن
الوزن			م	ف	ا	ع	ل		
الجذر				ر		ب	ط		

خامساً: نبحث عن الجذر المستخرج (ربط) في قاعدة بيانات الجذور .

- إذا موجود اذاً الكلمة صحيحة
- إذا غير موجود اذاً هنالك خطأ إملائي

أنظر الشكل التالي لشاشات التحليل:

مثال: مكتباتهم الكبيرة

شكل () أدخال الكلمات المراد تحليلها

شكل () نتائج التحليل

(١٠) أشهر برامج التحليل الصرفي[٦]:

ومن أهم الشـــركات العاملة في المجال هي التي توفر التدقيق الإملائي والنحوي، مثل صـــخر http://www.sakhr.com ، وكولتك http://www.coltec.net ، و آي بي إم www.ibm.com. وهناك ايضا شركة زيروكس http://www.xerox.com ، المشهورة في مجال إنتاج الآلات الناسخة والطابعات، والتي طورت أبحاثاً متطورة للتحليل الصرفي، من خلال مركزها للأبحاث في أوربا http://www.xrce.xerox.com/ ، وشـــركة عالمية أخرى تهتم باللغة العربية، وهي شركة L&H http://www.lhsl.com ، ولكنها متخصصة أصلاً بتقنيات معالجة اللغات الطبيعية، وتطور بعض الشـركات الأخرى محللات صـرفية بسـيطة، تسـتخدمها داخلياً في تطوير منتجات وتقنيات أخرى خاصة بها، مثل شركة نولدج فيو.

محلل شـركة زيروكس: يقبل المحلل الكلمات العربية المشـكلة كلياً أو جزئياً، ويعيد عدداً من الحلول الممكنة، ويقل عدد تلك الحلول كلما اســـتخدمت علامات التشـــكيل في الكلمات المدخلة. يقدم كل حل جذر الكلمة وتفعليتها وأي سـوابق أو لواحق ممكنة، وتسـتطيع اختبار هذا المحلل عبر إنترنت بزيارة الصفحة

المنتجات المعروضة http://www.xrce.xerox.com/research/mltt/arabic/

١) برنامج ArabMorpho

يمثل هذا البرنامج المحرك الأســاســي للتحليل الصــرفي العربي والوســم وأجزاء الكلام والتحليل الدلالي المعجمي باســـتخدام المقاطع، الأمر الذي يعزز من مرونته ونطاق التغطية فيما بعد

٩٩,٨% ويميزه عن البرامج المنافسة التي تحلل الوحدات الصرفية والمعجمية باستخدام ذخيرة الألفاظ، وفي حالة استنفاد القواعد الصرفية، يتم تطبيق تحليل إحصائي ديناميكي من أجل فك غموض الألفاظ بنسبة دقة تصل إلى ٩٦% وهو من إنتاج أروى أي.

٢) برنامج صخر للتحليل الصرفي[٧]

يتيح المعالج الصرفي المتعدد الأطوار أو Multi-Mode Morphological Processor من صخر المعالجة العميقة للكلمة العربية المفردة، ويغطي هذا المحلل نطاق الكلمات العربية بالكامل؛ الحديث منها والقديم.

يقوم المحلل بالتعرف على جميع أشكال جذر كلمة، أي أنه يقوم باستخلاص أصل الكلمة بعد تجريدها من اللواصق، ولا يتوقف عند هذا الحد بل يتخطى ذلك لاستخلاص البيانات الصرفية للكلمة مثل الجذر والميزان الصرفي لها (Morphological Pattern) وقسم الكلم الخاص بها.

يعمل المركب بوضع عكسي، حيث يعيد توليد الكلمة من أشكالها الصرفية (الساق والجذر والميزان الصرفي لها وقسم الكلم الخاص بها واللواصق).

قامت صخر بتصميم المعالج الصرفي لإجراء تحليل على مستوى الكلمة عن طريق إعادة التوليد والمقارنة. وبعبارة أخرى، يمكننا اعتبار التحليل صحيحًا إذا كان المعالج الصرفي قادرًا على إعادة إنشاء ومقارنة الناتج بالكلمة الأصلية." ويعاب على هذه المحللات كبر مساحة البيانات المخزنة بها وبطئها عند التنفيذ وهذه العيوب قد تتسبب في عدم تمكن بعض التطبيقات الأخرى من الاستفادة منها لعدم تمكن هذه التطبيقات من دمجها في برامجها؛ حيث سيصبح حجم هذه البرامج كبيرا جدا..

كما يلاحظ وجود بعض القصور أثناء تنفيذ عملية التحليل فمثلا: إذا قمنا باستخدام برنامج صخر للتحليل الصرفي فإن مفردات مثل: (اضتغط – ييجل – ياجل – ضَفْطً – ضفاطة – ضافط – تطلية – أظآر – أظؤر – متئب – إجْدَان – وَجِيجُ – إيجاح – أَوْجَحَ – يُوجِحُ) لن يمكننا العثور على نتيجة لها.

٣) منتج شركة صخر Sakhr

أنتجت الشركة برنامج يعمل كمحلل صرفي يقوم بتحديد الجذر والجذع والقالب الصرفي morphological pattern والنوع الصرفي للكلمة Part of speech POS. والبرنامج أيضاص يعمل في اتجاه معاكس كمولد صرفي يولد الكلمة من الجذر والجذع والقالب الصرفي morphological pattern والنوع الصرفي للكلمة Part of speech POS. شركة صخر أنتجت محللا صرفيا والذي يعد معالجا صرفيا لصيغ متعددة Multi-Mode Morphological Processor (MMMP). وهو يغطي اللغة العربية المعاصرة والتقليدية modern and classical Arabic. ويحدد الصيغة الأساسية برفع الزوائد.

٤) منتج شـــركـة اكســـروكس Xerox، محلـل ومولـد بيســـليز أكسوركس Beesley's Xerox:

بيسلي [8] طور مولد ومحلل صرفي لاستخدامه كمساعد في التعليم وكجزء من معالجة اللغات الطبيعية وهو متاح على الشبكة على [9]. وهو محلل صرفي من مستويين: الأول للجذور roots والقوالـب patterns والثـاني للزوائـد وحروف الجر وأدوات التعريف وأدوات الربط التي ترتبط بالكلمة. ويســـتخدم قاموس يحتوي على 4930 جذر بينما تكون الســـوابق واللواحق والجذر 77,800 صـورة. هذا المحلل للكلمات المشكلة تشكيلاً كاملاً أو تشكيل جزئي أو غير مشكلة. وهو يعطي جميع الاحتمالات بالتشـــكيل. فمثلاً كلمة "درس" تعطي ١٠ صـــور. و"كتب" تعطي ١٥ صورة. ولكننا لاحظنا أن ترتيب الصور لم يأخذ في الاعتبار التكرار الإحصائي. وهو عبارة عن محلل صـــرفي تم بناءه على أســـاس تقنية finite state. هذا المحلل يقبل كلمات العربية الفصحى ويحللها تحليلاً صرفياً بالإضافة للتفسير بالإنجليزية.

المحللات المتاحة على الشبكات

٥) محلل باك ولتر للغة العربية Buckwalter

برنامج باك وولتر للتحليل الصـــرفي يســـتخدم من قبل الاتحاد الألماني للبيانات اللغوية لتدوين النصـوص العربية، ويحتوي على ٧٧٨٠٠ جذر تمثل ٤٥ ألف مدخل معجمي ويسـتخدم المحلل لأغراض الترجمـة، لكنـه يعـاني من خلط العربيـة بـالرومـانيـة ولا يتعرف على علامـات الترقيم العربية. كتب محلل باك ولتر بلغة البرمجة برل Perl. يستخدم من قبل الاتحاد الألماني للبيانات اللغوية لتدوين النصـــوص العربية. ويحتوي 77,800 جذر تمثل 45,000 مدخل معجمي [١٠] ويستخدم المحلل لأغراض الترجمة. وقد نقده بيسلي [٩] بأنه:

- ليس قياسيا.
- صعوبة خلط العربية بالرومانية.
- و لا يتعرف علامات الترقيم العربية لأنها تمثل بحروف.

هو متاح على [١٠].

يقوم البرنامج بقبول الكلمة باللغة العربية وتحويلها للمقابل اللاتيني وبجزئها للســـوابق واللواحق والجذع والتحقق منهم في قاميســـه الخاصـــة. وإذا وجودوا في القواميس يطبع النتائج في ملف المخرجات output file. ومشـكلة هذه المخرجات أنها باللغة اللاتينية. وقواميس باك ولتر عباره عن ثلاثة معاجم ؛ سوابق ولواحق وجذوع.

- معجم السوابق **Prefixes dictionary**

يحتوى كل ســـوابق الكلمة العربية وهم: و ، ف ، ب ، ك ، أ ، ي ، س ، ل ، ت ، ن وليس ال.

- معجم اللواحق **Suffixes dictionary**

يحتوى كل لواحق الكلمة العربية وهم: ا, ت ، ه ، أ ، ك ، ن ، ف ، ي ، و **وليس ة.**

- معجم الجذوع **Stems dictionary**

يحتوي كل جذوع الكلمات العربية.

أقسام الكلمات Categories

الإسم

الفعل التام

الفعل الناقص

١. **الأحرف**character :

الروابط conjunctions - **حروف الجر** preposition– **أدوات التوكيـد** emphatic particle

٢. **الصفات** adjective

٣. **الظروف** adverb

٤. **الضمائر** pronoun

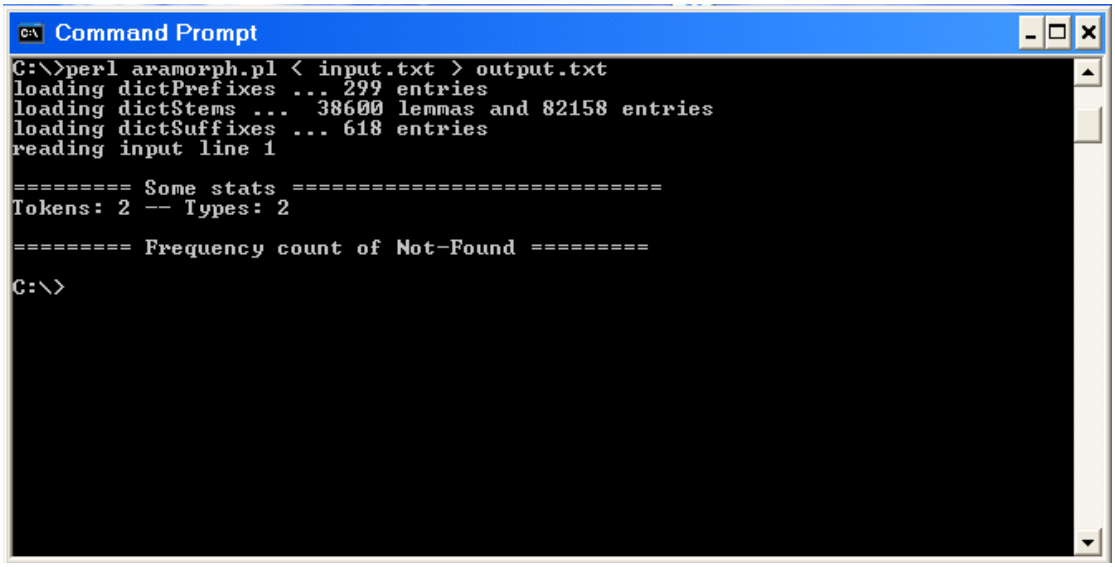
ويعمل بالأوامر كالتالي:

```
C:\ perl    aramorph.pl < input.txt > output.txt
```

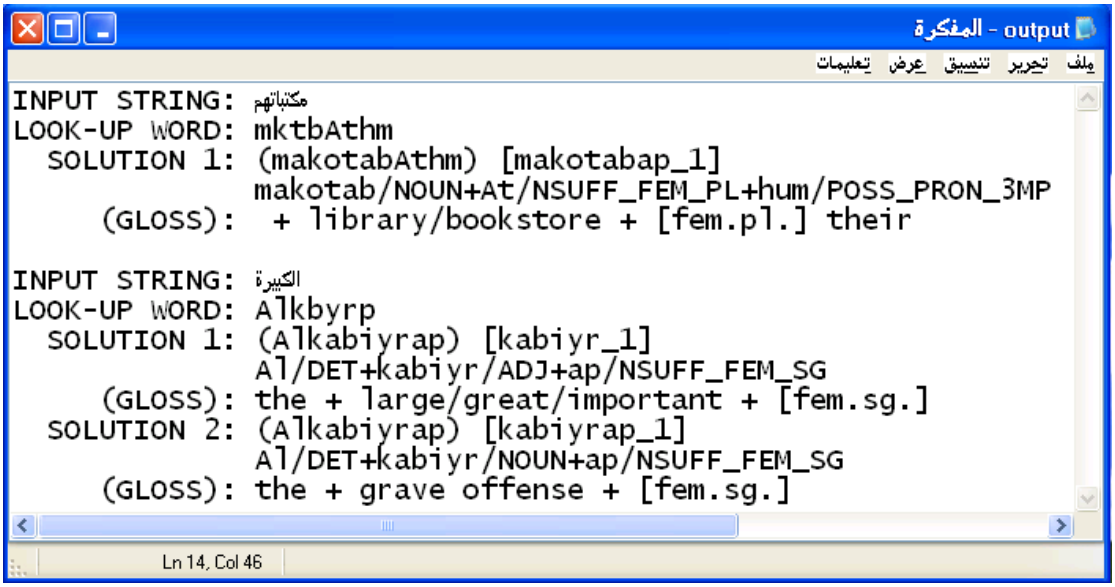
وملف الإدخال: Input.txt → Arabic text encoding CP-1256

وملف الإخراج : Output.txt → the result of every word in English

شكل() ملف المدخلات



شكل () التنفيذ



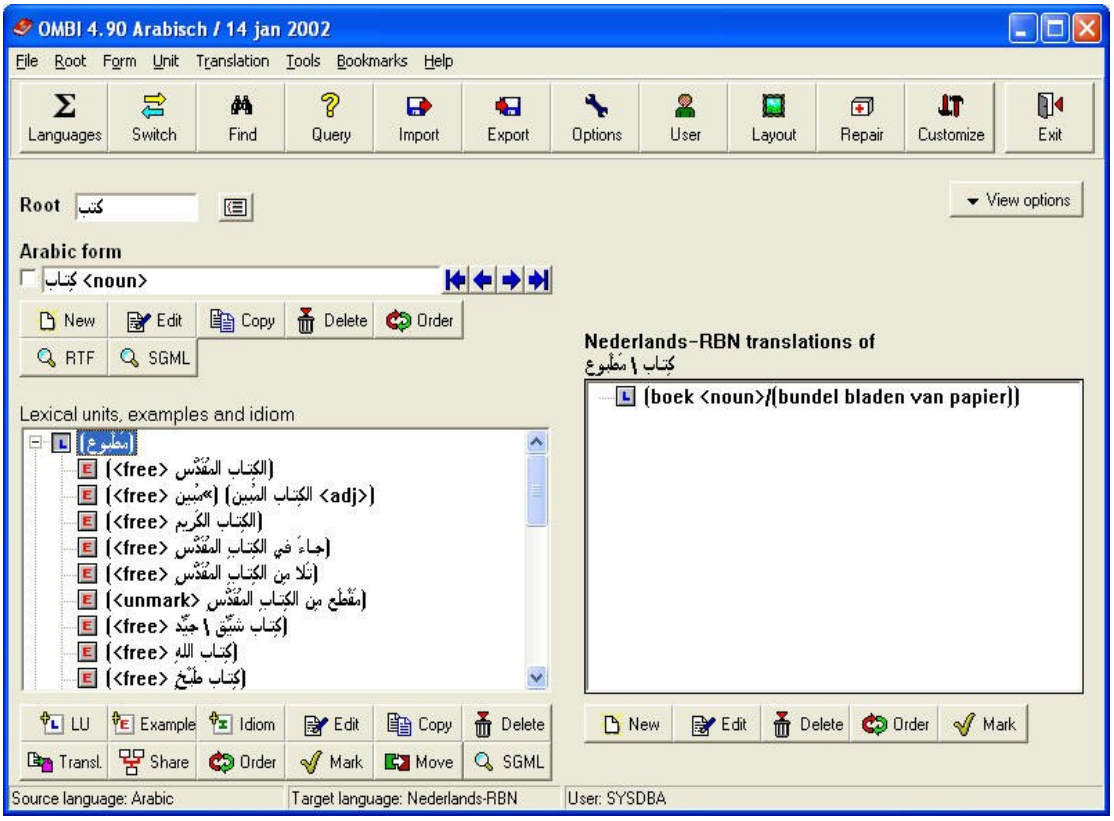
شكل () المخرجات

٦) المُحلِّل الصَّرفي لِمُدَوَّنة جامِعة نيجمجن

NIJMEGEN Corpus Morphological Analyzer

- اسم البَرنامَج: Omkeebaar Bilinguaal Bestand.
- الجِهة المُنتِجة: جامِعة نيجمِجن بهولندا NIJMEGEN University.
- الغرض مِن البَرنامَج: التَّحليل الصَّـــرفِيّ لِلمُدَوَّنة العربِيَّة المُنجَزة بالجامعة الهولندِيَّة NIJMEGEN. (راجع الفصل الأوَّل: المُدَوَّنات العربِيَّة المُنجَزة إلكترونِيًّا).
- الأنظِمة القابِلة لِعمل البرنامَج: أنظِمة تشغيل Microsoft.
- صُمِّمَ البَرنامَج باستِخدام لُغة Delphi.
- معلومات البَرنامَج على الشَّبكة العنكبوتِيَّة:

http://www.let.kun.nl/wba/Content1/1.4.2%20ombi.htm

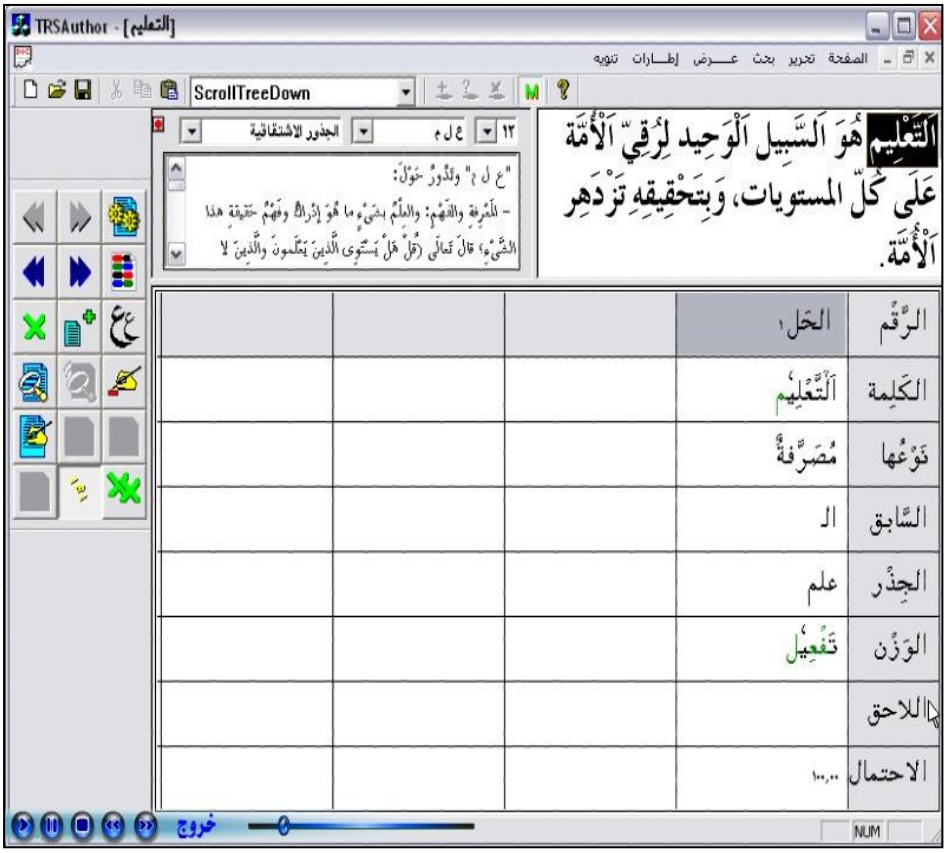


شكل (٧) المُحلِّل الصَّرفِي لجامِعة NIJMEGEN

٧) المُحلِّـل الصَّـــرفي لشـــركـة RDI RDI's Morphological Analyser

- اسم البَرنامَج: RDI ArabMorpho.
- الجِهة المُنتِجة: الشركة الهندسيَّة المِصريَّة لتطوير نُظُم المعلومات RDI.
- الغرض مِن البَرنامَج: التَّحليل الصَّرفِيّ العَميق لِمُفرَدات اللُّغَة العَرَبِيَّة.
- الأنظِمة القابِلة لِعمل البرنامَج: أنظِمة تشغيل Microsoft.
- معلومات البَرنامَج على الشَّبكة العنكبوتِيَّة:

http://www.rdi-eg.com/rdi/Technologies/Arabic%20NLP_m.htm#ArabMorpho



شكل رقم (٨) المُحلِّل الصَّرفِي لشركة RDI المِصريَّة

٨) **مجـذع كريم /Kareem Stemmer Darwish's morphological analyzer:** اسـتطاع درويش [١١] تطوير محلل صـرفي للغة العربية ALPNET وهو غير مجدي 'shallow'إذ يعطي فقط الجذور المحتملة possible roots الكلمات. ولهذا يمكن اعتباره مجزئ كلمات stemmer فقط. ويتكون من معالجين:

١. النموذج البنائي 'Build-Model': ويسـتخدم قائمة من أزواج جذور الكلمات كمدخل ليستخرج قائمة للسوابق واللواحق لحساب ترجيح likelihood ظهور السوابق واللواحق. هذه القائمة (279,606 كلمة) تنتج آليا باستخدام المحلل الصرفي.

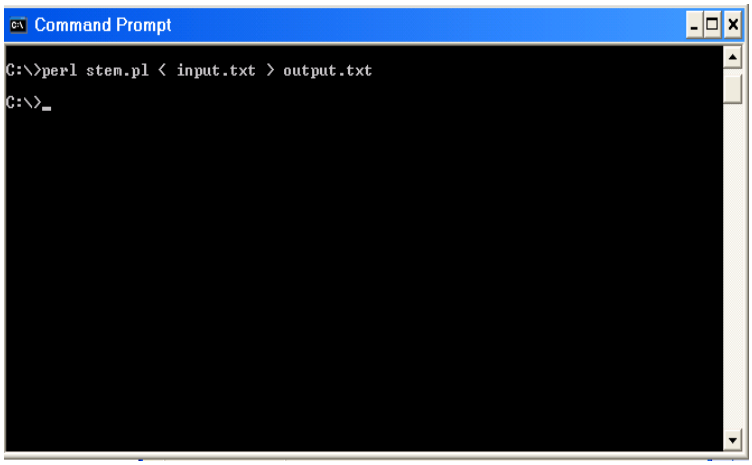
٢. النموذج الكشـفي عن الجذر 'Detect Root': يحسـب احتمال الجذور والزوائد كتركيبة واحدة one combination. والمشـكلة هي ضـخامة عدد الكلمات المولدة over-generate وإنتاج كلمات لا وجود لها. وتحل هذه المشكلة بالبحث في قائمة القاموس او التحقق منها يدوياً. وسـرعته 40,000 كلمة / دقيقة على ماكينة بنتيم. وتقول لطيفة أنه وبالرغم من إتاحته على الشـبكة لم نتمكن من إنزاله. وقد **كتب بلغة برل. يقوم بحذف**

سـوابق ولواحق الكلمة. ومن عيوبه أن المخلات والمخرجات باللغة اللاتينية. ويعمل بالأوامر التالية

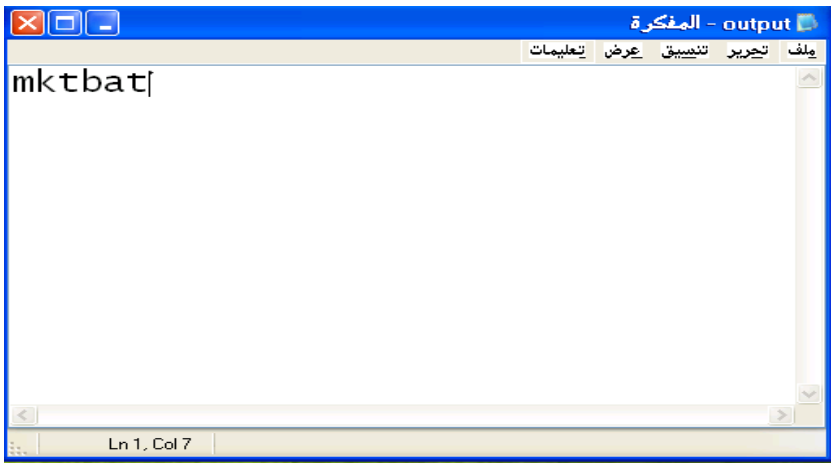
```
C:\ perl   stem.pl < input.txt > output.txt
```

Input.txt → كلمات عربية بحرو إنجليزية بكود UTF-8 or CP-1256

Output.txt → بحروف لاتينية



شكل(٩) تنفيذ البرنامج

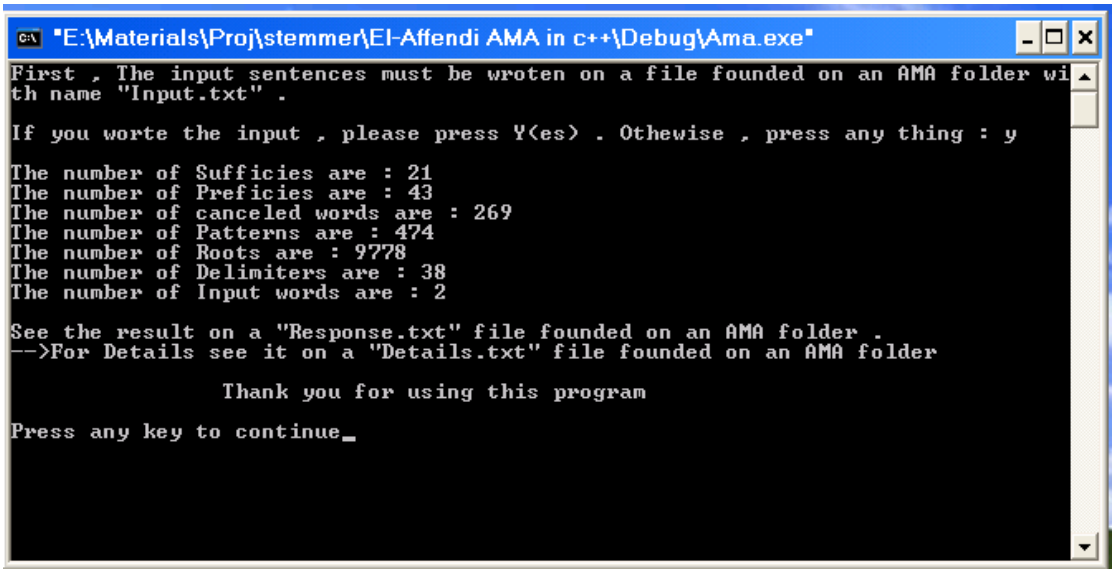


شكل(١٠) ملف المخرجات

٩) الأفندي[١٢]:

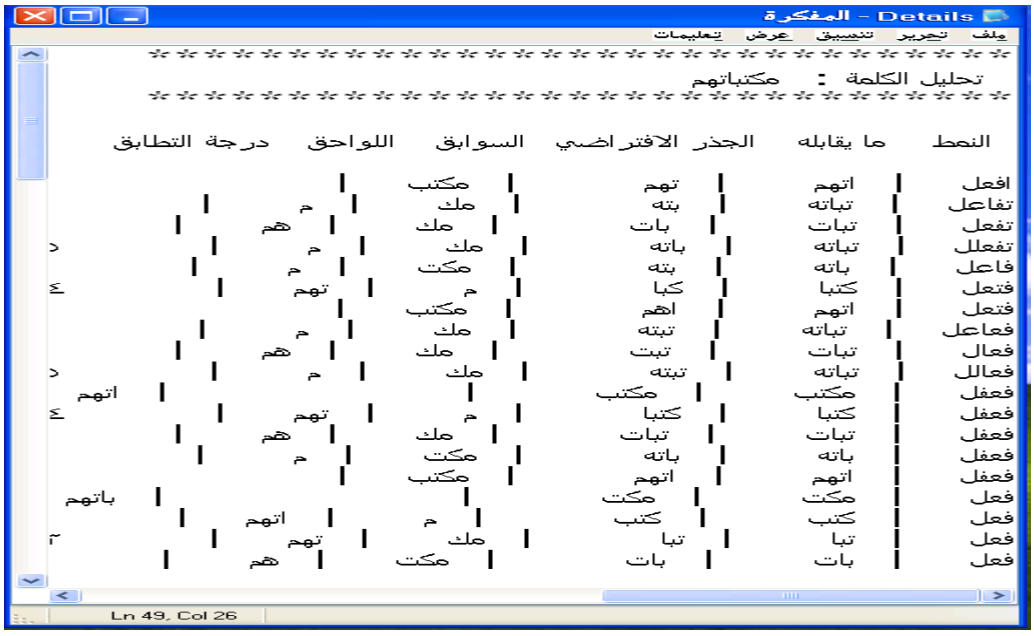
عبارة عن رسـالة دكتوراة بلغة سـي++. يأخذ الكلمة كمدخل ويخرج ملف بتفاصـيلها وعيبه أن يعمل على الجذر لا الجذع. بالنسبة للمدخل له نفس المدخل السابق.

أما المخرج انظر الشكل التالي:

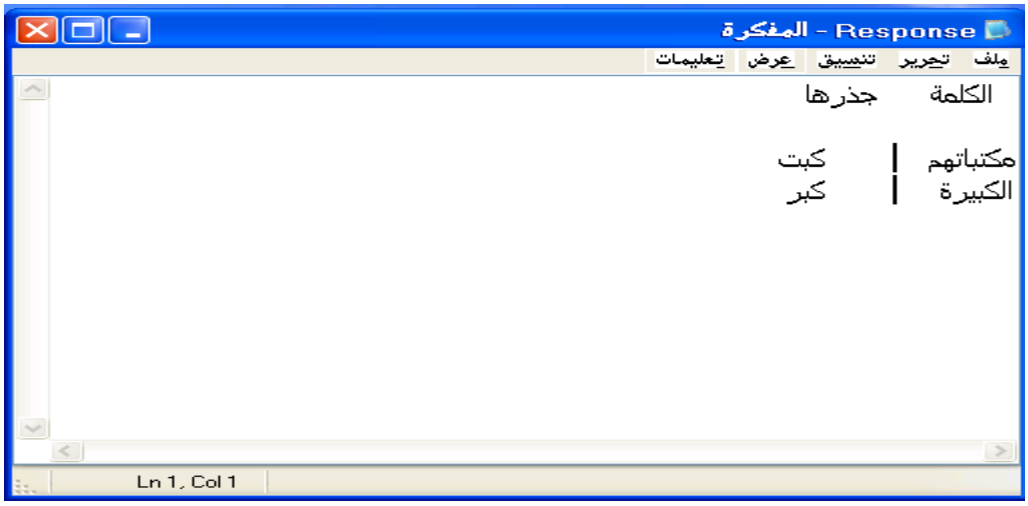


شكل (١١)مخرج البرنامج احصاء تحليل الكلمة

أما ملف تفاصيل التحليل



شكل(١٢) ملف تفاصيل التحليل



شكل(١٣) مخرج لتوضيح المبني على الجذر

١٠) **مجذع كوجا Khoja Arabic Stemmer لشيرين خوجة**

١١) **نظام ليه لاركي[١٣]** انظر فقرة أنظمة الإزالة.

١٢) **النظام Restrict [2]**

طريقة العمل

- توحيد الاخطاء الاملائية كما هو موجود في الانظمة السابقة مع اضافة
  - استبدال حرفي "و"، "ء" بحرف "ؤ"
  - استبدال حرفي "ي"، "ء" بحرف "ئ"
  - استبدال حرفي "ى"، "ء" بحرف "ئ"
  - استبدال حرفي "ا" ، "ا" بحرف "ا"
  - فصل الكلمة الى اثنتين اذا وجد في وسطها "ة"
  - فصل الكلمة الى اثنتين اذا وجد حرف لا تتغير صورته اذا اتي بعده حرف اخر متبوعا بـ "ال" على ان تكون تحتوي كلتا الكلمتين على اربعة احرف او اكثر
- ازالة اللواحق من اول الكلمة
  - ازالة أحرف العطف والجر
    - تكرار الحرف الاول اذا كان "ف"، "ب"، "ك"، "ل"، او "و"
    - ازالته اذا كانت الكلمة الجديدة موجدة بالمعجم
    - اذا كان الحرفان الاولان "للـ" استبدلهما بحرف "لـ" كذلك اضف حرف الالف الى الكلمة قبل

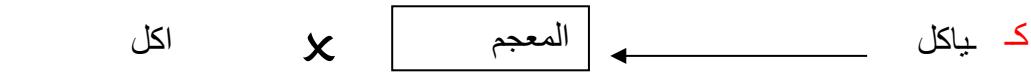
- اذا كانت الكلمتان الجديدتان صحيحتان احذف اللام الاولى

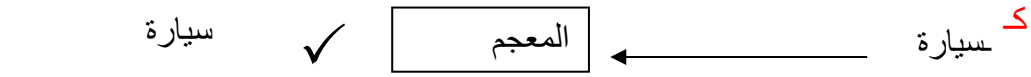
مثال

ازالة حرف الوال من كلمة وليمة

- **إزالة اداة التعريف "الـ"**
- **ازالة لواحق الافعال**

- يتم استخدام حرف جر الى بداية الكلمة للتاكد من انها فعلا ام اسما
- اذا لم تقبل الكلمة حرف الجر (فعلا) يتم ازالة الحرف الاول
- بالنسبة للكلمات التي تبدا بحرف السين "ســـ" يتم التاكد اولا ان الكلمة ليست اسما وذلك باضافة اداة التعريف "الـــ" واستخدام حرف الكاف كذلك.
- قبلت الكلمة إحدى هاتين اللاحقتين لا يتم ازات حرف السين
- اذا لم تقبلهما يتم ازالته وما يتبعه من احرف المضارعة.





- **ازالة اللواحق التي تلحق آخر الكلمة**

- يتم حذف الضمائر "ها"، "هم"، "هما"، "هن" من اخر الكلمة
- يتم حذف اللاحقة "ان" فقط بعد التأكد من أنهافعلا لاحقة المثني وذلك باســـتبدالها باللاحقة "ين" واختبار وجود الكلمة الجديدة في المعجم.
- يتم استبدال اللاحقة "ات" الى "ة" اذا نتج عن هذا كلمة موجودة بالمعجم.
- يتم حذف لاحقة المذكر السالم "ون" اذا نتج عن استبدالها بـ "ين" كلمة موجودة بالمعجم.
- يتم حذف اللاحقة "ين" اذا نتج عن استبدالها بـ"ون" او "ان" كلمة موجودة بالمعجم

• يتم حذف اللواحق "وا"، "ية"، "يه"، "ة"، "ه"

• يتم ازالة اللاحقة "ي" من اخر الكلمة اذا نتج عن حذفها كلمة صـــحيحة، او كان ناتج استبدالها بـ "ها" و"ه" كلمتان صحيحتان

– بعد عملية حذف اللواحق يتم ارجاع بعض الاحرف الى اصلها

تقلب التاء المفتوحة تاء مربوطة اذا كانت الكلمة الجديدة موجدة في المعجم، إذا ســـبق التاء حرف الف يقلبا معا الى تاء مربوطة واعادة عملية ازالة اللواحق من جديد.

مثال:

"مدرستها" بعد ازالة "ها" تصبح "مدرست" تقلب الى "مدرسة"

"سياراتهم" بعد ازالة "هم" تصبح "سيارات" تقلب الى "سيارة"

– يقلب حرف "ؤ" و "ئ" الى همزة "ء"

مثال "سماؤهم" بعد ازالة "هم" تصبح "سماؤ" تقلب الى "سماء"

"مسائي" بعد ازالة "ي" تصبح "مسائ" تقلب الى "مساء".

١٣) برنامج تحليل صرفي للكلمات العربية: انظر فهرس **(٢)**.

١٤) نظام تحليل صرفي للغة العربية[١٥]:

يمكن أن يولد الأفعال والمشـــتقات الإســـمية وتصـــريفاتها للجذور الثلاثية والرباعية معتمدا على قواعد الصـــرف العربي ومســـتخدما قاعدة بيانات النظام. لم نســـتطع استخدامه. وهو مفتوح المصدر على[١٦].

فهرس نموذج تفصيلي لأحد المحللات للغة العربية

المُحلِّل الصَّرفي للمُعجَميّ Tim Buckwalter

Buckwalter Arabic Morphological Analyser

- اسم البَرنامَج: Buckwalter Arabic Morphological Analyser.
- الجِهة المُنتِجة: Tim Buckwalter، جامِعة بنسلفانيا.
- الغرض مِن البَرنامَج: المعالجة الآلِيَّة لِنُصُـــوص المُدَوَّنات العربِيَّة، ويعمَل – كذلك – كمُعَنوِنٍ للتَّراكيب العَرَبِيَّة.
- الأنظِمة القابِلة لِعمل البرنامَج: أنظِمة تشغيل Microsoft.
- معلومات البَرنامَج على الشَّبكة العنكبوتِيَّة:

http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49

نموذَج لِمادَّةٍ تَمَّ تحليلُها صَرفِيًّا باستِخدام المُحلِّل الصَّرفِيّ للمُعجَمِيّ

Tim Buckwalter

المَرتَبة المُقَعَّرة/ يُوسُف إدريس

المرتبة المقعرة

في ليلة الدخلة والمرتبة جديدة وعالية ومنفوشة رقد فوقها بجسده الفارع الضخم واستراح إلى نعومتها وفخامتها وقال لزوجته التي كانت واقفة إذ ذاك بجوار النافذة انظري هل تغيرت الدنيا ونظرت الزوجة من النافذة ثم قالت لا لم تتغير

INPUT STRING: المرتبة
LOOK-UP WORD: Al_m_rt_b_p
  Comment: Al_m_rt_b_p NOT FOUND
 ALTERNATIVE: Al_m_rt_b_h
  Comment: Al_m_rt_b_h NOT FOUND

INPUT STRING: المقعرة
LOOK-UP WORD: Al_m_q_E_rp
  Comment: Al_m_q_E_rp NOT FOUND
 ALTERNATIVE: Al_m_q_E_rh
  Comment: Al_m_q_E_rh NOT FOUND
NEWLINE
INPUT STRING: في
LOOK-UP WORD: fy
 SOLUTION 1: (fiy) [fiy_1] fiy/PREP
  (GLOSS): + in +
 SOLUTION 2: (fiy~a) [fiy_1] fiy/PREP+~a/PRON_1S
  (GLOSS): + in + me
 SOLUTION 3: (fiy) [fiy_2] Viy/ABBREV
  (GLOSS): + V. +

INPUT STRING: ليلة
LOOK-UP WORD: lylp
 SOLUTION 1: (layolap) [layolap_1] layol/NOUN+ap/NSUFF_FEM_SG
  (GLOSS): + night/evening + [fem.sg.]

فـــهـــرس شـــــكـــل مـــخـــرج بـــرنـــامـــج Arabic Morphological Analyzer

برنامج تحليل صرفي للكلمات العربية[١٤]

ندخل الجذر الثلاثي أو الرباعي للكلمة وتحرج المشتقات بالشكل التالي:

شكل لمشتقات الجذر"حرم" مجموعة مختارة من المخرج

52 Possible Stems for the Root: حرم **Hrm**			
Nr	**Stem**	**Category**	**Gloss**
1	**Harum-** حَرُم	perfect verb stem + PV_intr inflections	**be forbidden**
2	**-Hrum-** حْرُم	imperfect verb stem + IV_intr inflections	**be forbidden**
38	**maHârim-** مَحارِم	noun stem + Ndip inflections	**forbidden/sacred**
39	**muHâramä-** مُحارَمَة	noun stem + N0 inflections	**Muharama**
40	**maHram-** مَحْرَم	noun stem + Nap inflections	**handkerchief**
41	**maHârim-** مَحارِم	noun stem + Ndip inflections	**handkerchiefs**
42	**taHrîm-** تَحْرِيم	noun stem + NduAt inflections	**prohibition/ban**
43	**'îHrâm-** إِحْرام	noun stem + NduAt inflections	**state of ritual consecration/pilgrimage garments**
44	**'îHrâm-** إِحْرام	noun stem + NduAt inflections	**state of ritual consecration/pilgrimage garments**
45	**iHtirâm-** اِحْتِرام	noun stem + NduAt inflections	**respect/honoring**
46	**iHtirâm-** اِحْتِرام	noun stem + NduAt inflections	**respect/honoring**
47	**iHtirâm-** اِحْتِرام	adverb stem + NF inflections	**out of respect for/in honor of**
48	**iHtirâm-** اِحْتِرام	adverb stem + NF inflections	**out of respect for/in honor of**
49	**maHrûm-** مَحْرُوم	adjective stem + Nall inflections	**deprived**
50	**muHarram-** مُحَرَّم	adjective stem + Nall inflections	**forbidden**
51	**muHarram-** مُحَرَّم	adjective stem + Ndip inflections	**Muharram (month)**
52	**muHtaram-** مُحْتَرَم	adjective stem + Nall inflections	**esteemed/respected**

فهرس بعض من نتائج تحليل [١٧] الموجودة على الموقع

```
Allyl   layol POS:N Al+ +SG +MASC
Allylp  layolap POS:N +FEM Al+ +SG
Allyly  layoliy~ POS:AJ Al+ +SG +MASC
Allylyp layoliy~ POS:AJ +FEM Al+ +SG
AlfA}z  fA}iz POS:N Al+ +SG +MASC, fA}iz POS:PN Al+ +SG +MASC
AlfA}zp fA}iz POS:N +FEM Al+ +SG
AlfA}ztyn       fA}iz POS:N +FEM Al+ +DU +ACCGEN
AlfA}zwn        fA}iz POS:N +MASC Al+ +PL +NOM
brDA    riDA POS:PN b+ +SG +MASC, riDaY POS:N b+ +SG +MASC
b>rD    >aroD POS:N b+ +SG +MASC
brd     barad POS:N +SG +MASC, bar~ad POS:V +PV +S:3MS, barad POS:V
+PV +S:3MS, barod POS:N +SG +MASC, barud POS:V +PV +S:3MS, burod
POS:N +SG +MASC
brdE    radoE POS:N b+ +SG +MASC
b>Hmd   >aHomad POS:N b+ +SG +MASC, >aHomad POS:PN b+ +SG +MASC
bHmd    Hamod POS:N b+ +SG +MASC, Hamod POS:PN b+ +SG +MASC
bHmdwn  Hamoduwn POS:PN b+ +SG +MASC
bhmhA   bahiym POS:N +PL +P:3FS, ham~ POS:N +P:3FS b+ +SG +MASC
AldyAn  day~An POS:N Al+ +SG +MASC, day~An POS:PN Al+ +SG +MASC
Al>dyAn diyn POS:N Al+ +PL
AldyAnp day~An POS:N +FEM Al+ +SG, diyAnap POS:N +FEM Al+ +SG
AldyAntyn       day~An POS:N +FEM Al+ +DU +ACCGEN, diyAnap POS:N +FEM
Al+ +DU +ACCGEN
mnZmty  munaZ~amap POS:N +FEM +DU +ACCGEN +POSS, munaZ~amap POS:N
+FEM +DU +ACCGEN +POSS +P:1S, munaZ~amap POS:N +FEM +SG +P:1S,
munaZ~am POS:AJ +FEM +DU +ACCGEN +POSS, munaZ~am POS:AJ +FEM +DU
+ACCGEN +POSS +P:1S, munaZ~am POS:AJ +FEM +SG +P:1S, munaZ~im POS:N
+FEM +DU +ACCGEN +POSS, munaZ~im POS:N +FEM +DU +ACCGEN +POSS +P:1S,
munaZ~im POS:N +FEM +SG +P:1S
mnZmtyn munaZ~amap POS:N +FEM +DU +ACCGEN, munaZ~am POS:AJ +FEM +DU
+ACCGEN, munaZ~im POS:N +FEM +DU +ACCGEN
mnZmwhA munaZ~am POS:AJ +MASC +PL +NOM +P:3FS, munaZ~im POS:N +MASC
+PL +NOM +P:3FS
mnZmw   munaZ~am POS:AJ +MASC +PL +NOM +POSS, munaZ~im POS:N +MASC
+PL +NOM +POSS
mHTtyn  maHaT~ap POS:N +FEM +DU +ACCGEN
mHtwAhA muHotawaY POS:N +P:3FS +SG +MASC
mHtwAh  muHotawaY POS:N +P:3MS +SG +MASC
mHtwmp  maHotuwm POS:AJ +FEM +SG
bArtkAbhA       AirotikAb POS:N +P:3FS b+ +SG +MASC
bArtkAbh        AirotikAb POS:N +P:3MS b+ +SG +MASC
bArtnrz bArtonraz POS:PN +SG +MASC
bA$rTp  $ariyT POS:N +PL b+
yrtDy   AirotaDaY POS:V +IV +S:3MS
yrtdyAn AirotadaY POS:V +IV MOOD:I +S:3MD
yrtDyhA AirotaDaY POS:V +IV +S:3MS +O:3FS
yrtdyh  AirotadaY POS:V +IV +S:3MS +O:3MS
>nATh   >anAT POS:V +PV +S:3MS +O:3MS
nATHAt  nATiHap POS:N +FEM +PL
n$AThm  na$AT POS:N +P:3MP +SG +MASC, na$iyT POS:N +PL +P:3MP
n$ATh   na$AT POS:N +P:3MS +SG +MASC, na$iyT POS:N +PL +P:3MS
bmHmwd  maHomuwd POS:N b+ +SG +MASC, maHomuwd POS:PN b+ +SG +MASC
bmHmyp  maHomiy~ POS:AJ +FEM +SG b+
bmhndsy muhanodis POS:N +MASC +DU +ACCGEN +P:1S b+, muhanodis POS:N
+MASC +DU +ACCGEN +POSS b+, muhanodis POS:N +MASC +PL +ACCGEN +P:1S
b+, muhanodis POS:N +MASC +PL +ACCGEN +POSS b+, muhanodis POS:N +P:1S
b+ +SG +MASC
```

الخلاصة:

استعرض التقرير ١٦ من المحللات التي تعمل على اللغة العربية وللآسف أكثرها تم بجامعات أو مراكز بحثية خارج المنطقة العربية وبتمويل اجنبي. بالنسبة للمحللات داخل المنطقة العربية مازالت في طور نمو وبعضها مبشر وذلك لنقص الدعم وعدم وجود منهجية تعاون بين المتخصصين. وبالنسبة للمحللات خارج المنطقة العربية أغلبها مخرجاته باللغة اللاتينية ويتعامل مع اللغة شكلياً. وأعلب هذه المحللات لم تغط جميع صور المشتقات بأنواعها المختلفة. وأغلب المحللات غير متاحة مجاناً أيضاً. إذن المطلوب منتج عربي يتعامل مع اللغة بتفهم ويوفر مخرج يسهل استعماله للعوام من ناحية وللمتخصصين من ناحية أخرى. ويفضل أن يتاح مجانا مفتوح المصدر قابل للتعديل والتطوير ليسمح باحتواء اللغة العربية ومشتقاتها بكافة الصور.

المراجع

(١) حيدر العميد، شيخة الكتبي، آمنة الكعبي، خديجة الشبلي، نيلا الشامسي، نورة النعيمي، شيخة المهيري ،"خطوات عملية باتجاه تطبيق مفهوم البحث بالمفاهيم لمحركات البحث والاسترجاع العربية"،

http://www.phillips-publishing.com/files/__v1.ppt

(٢) عبدوسالم .ف. أ. نوسري، " معالجة النصوص العربية لأجل الفهرسة والإسترجاع"

http://www.iera.ac.ma/iera/CITALA07/archivecitala07/Presnetation/nzesr_CITALA07.ppt

(٣) عبد الله بن عبد الرحمن الزامل، «العلاقة الصرفية بين الجذور والأوزان (تصنيف جديد لجذور اللغة العربية)،ندوة الحاسب و اللغة العربية، الرياض، المملكة العربية السعودية ٢٠٠٧م،

(٤) عبد الإله السفياني، المعالجة الآلية للغة العربية (التحليل الصرفي)

http://faculty.kfupm.edu.sa/ics/muhtaseb/Teaching/ACMorphoAnal.ppt

(5) Khoja, S. "APT: An automatic Arabic part-of-speech tagger." PhD thesis. University of Lancaster, 2003.

http://www.iscal.org.sa/

(٦) "تقنيات اللغة العربية، إلى أين؟" ،

http://www.ccse.kfupm.edu.sa/~husni/ICS484/WebPAges/Munawes/Aracbic.htm

(٧) مجلة لغة العصر عدد (٩٢) أغسطس ٢٠٠٨.

(8) Beesley, K. (2001). Finite-State morphological analysis and generation of Arabic at Xerox research: status and plans in 2001. In http://www.elsnet.org/acl2001-arabic.html.

(9) Beesley, K. (2003). Xerox Arabic Morphological Analyser Surface-Language [Online].

http://www.xrce.xerox.com/competencies/content-analysis/arabic-inxight/arabic-surf-langunicode.pdf.

(10) Buckwalter Documentation,

http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49

,Buckwalter Arabic Morphological Analyzer Version 1.0

http://home.byu.net/jmd56/download/buckwalter_morphan_1.tar.gz

(11) Darwish, K. (2002). Building a shallow morphological analyser in one day.

http://www.cs.umd.edu/Library/TRs/CS-TR-4326/CS-TR-4326.pdf

(١٢) الأفندي.م.أ عمل بحث صرفي على النت، جامعة الملك سعود الرياض، السعودية.

(13) Larkey, Leah S., Ballesteros, Lisa, and Connell, Margaret. (2002) Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis In Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002), Tampere, Finland.

(١٤) برنامج تحليل صرفي للكلمات العربية

http://lexanalysis.com/araflex/araflex.html

(١٥) محمد معموري وآخرين ،"نظام تحليل صـرفي للغة العربية"، مجلة علوم اللغة الماليزية،١٩٩٥،

http://myais.fsktm.um.edu.my/756/

(١٦) موقع صرف ،"نظام المحلل الصرفي"، مفتوح المصدر

http://sourceforge.net/project/shownotes.php?release_id=533655&group_id=202978

http://sourceforge.net/project/downloading.php?group_id=202978&use_mirror=switch&filename=sarf.jar&a=77932681

(17) Unsupervised Morpheme Analysis -- Morpho Challenge 2008, http://www.cis.hut.fi/morphochallenge2008/