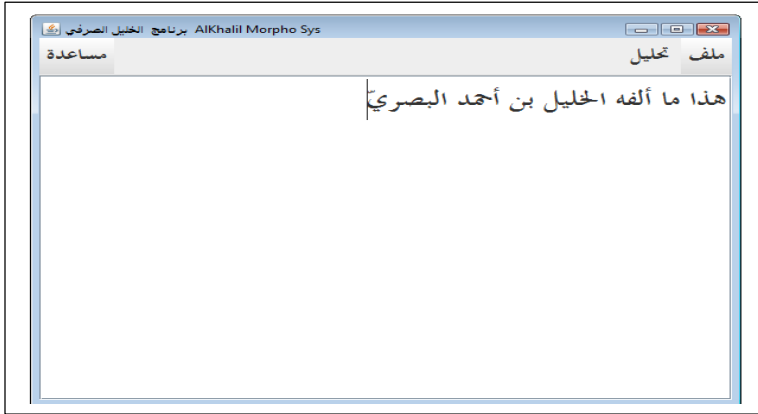




المنظمة العربية للتربية والثقافة والعلوم

تقرير مفصّل حول برنامج التحليل الصرفي في اللغة العربية -برنامج الخليل الصرفي-



الدخول INPUT	الكلمة المشكولة Voweled Word	السايق Prefix	نوع الكلمة Type	الوزن Pattern	الجذر Root	الحالة الإعرابية POS Tags	اللاحق Suffix
هذا	هَذَا	#	اسم إشارة	#	#	#	#
هذا	هَذَا	#	مصدر أصلي	فَعَا	هَذَا	مفرد مذكر منصوب لكرة	#
هذا	هَذَا	#	فعل ماضٍ مبني للمعلوم	فَعَا	هَذَا	ثلاثي مجرد مسند إلى العائنين (هما) متعَد ولازم	#
هذا	هَذَا	#	فعل أمر	فَعَا	هَذَا	ثلاثي مجرد مسند إلى المخاطبتين (أنتما) متعَد ولازم	#
هذا	هَذَا	#	فعل ماضٍ مبني للمجهول	فَعَا	هَذَا	ثلاثي مجرد مسند إلى العائنين (هما) متعَد ولازم	#
هذا	هَذَا	#	فعل ماضٍ مبني للمعلوم	فَعَا	هَذَا	ثلاثي مجرد مسند إلى الغائب (هو) متعَد ولازم	#
ها	هَا	#	اسم شرط	#	#	#	#
ها	هَا	#	اسم موصول	#	#	#	#

تقديم:

أنجز نظام الاشتقاق والتصريف، برنامج الخليل الصرفي، تحت مظلة المنظمة العربية للتربية والثقافة والعلوم خلال العام 2010، بالتعاون مع جامعة محمد الأول بالمملكة المغربية، ومعهد بحوث الحاسب في مدينة الملك عبد العزيز للعلوم والتقنية بالمملكة العربية السعودية. يرمي هذا النظام إلى تمكين الحاسوب من اشتقاق (توليد) الأفعال والأسماء المشتقة والمصادر، وتصريفها، انطلاقاً من جذورها الثلاثية أو الرباعية، بالاعتماد على قوانين النحو والصرف وعلى المعجم الحاسوبي لهذا النظام، وعلى قواعد معطيات خاصة، ومن أهم مزاياه أنه مفتوح المصدر، فبرنامجُه، وخوارزمياته، وقواعدُ معطياته، ووثائقُه، جميعُها متاحة. وبذلك يستطيع العاملون في مجال المعالجة الآلية للغة العربية المخولون، وفي مجال المعلوماتية تطوير البرنامج والخوارزميات، وتعديل قواعد المعطيات إضافةً وحقناً وتحديثاً.

الهدف من برنامج الخليل الصرفي ومزاياه:

الغرض من نظام الخليل الصرفي هو تحليل الكلمة لتحديد المعارف الصرفية المتعلقة بها كجذورها وسابقتها ولاحقتها ونوعها ووزنها...ويُعدُّ هذا النظام من الأدوات الرئيسية المفتوحة المصدر في معالجة الآلية للغة العربية، لذا فإنه يدخل في بنية الأنظمة الأخرى لمعالجة اللغة العربية، نذكر منها المعجم العربي الحاسوبي النفاعلي.

دَخَلَ نظام التحليل الصرفي كلمةً مفردة، أو مجموعةً كلماتٍ تَوَلَّفَ جملةً (أو نصًّا). فإذا كان الدخل جملةً (أو نصًّا)، عالَجَ النظامُ كلماتها تبعاً، وأعطى تحليلَ كلِّ منها على حدة مستقلةً عن سياق النص. يمتاز دخل النظام بأنه يقبل الكلمة مشكولةً كلياً، أو جزئياً، أو غير مشكولة البنية.

ومن مزايا نظام التحليل الصرفي أنه يراعي الوجوه المختلفة للكلمة؛ في حالتين:

الأولى: إذا كانت غير مشكولة (نحو: وهم = وَهْمٌ، وَهْمٌ، وَهْمٌ، وَهْمٌ، وَهْمٌ...)

الثانية: إذا كانت مستقلةً عن سياقها في النص؛ نحو:

فَرَمْتُ = فَ + رَمْتُ (من الفعل: رَمَى يَرْمِي)،

فَرَمْتُ (من الفعل: فَرَمَ يَفْرُمُ).

ويختلف عدد هذه الوجوه من كلمة إلى أخرى تبعاً لاستعمالات الكلمة. ولما كان نظام التحليل الصرفي يعالج الكلمة مستقلةً عن سياقها في النص، فإن نتيجة تحليله تتضمن جميع الوجوه الممكنة للكلمة. وتنتهي عند هذا الحد مهمة المحلل الصرفي، لتبدأ مهمة المحلل النحوي الذي يختار من بين هذه الوجوه ما يوافق سياق النص تماماً.

ومن خصائص الكلمة في اللغة العربية أيضاً أنها لا تعدو أن تكون فعلاً أو اسماً أو حرفاً؛

أما الفعل، فيستوعب جميع الأفعال الثلاثية والرباعية، المجردة والمزيدة؛ (نحو: رَفَع، نَقَّبَ، كَبَّكَبَ، تَسَرَّلَ...). وأما الاسم، فيشتمل على جميع الأسماء والمصادر والصفات والظروف... (نحو: كتاب، عالم، كبير، وراثته، أمس...). وأما الحرف، فيتضمن جميع حروف الجر والنصب والجزم والعطف... (نحو: من، أن، لا، و...).

وبسبب غياب الشكل، يمكن أن تشترك وجوه التحليل الصرفي للكلمة في:

- الإسمية والفعلية؛ نحو: ورد = وَرَدَ (اسم)، وَرَدَ (فعل).

- أو الاسمىة والحرفىة؛ نحو: رب = رَبُّ (اسم)، رَبُّ (حرف).
- أو الفعلىة والحرفىة؛ نحو: عدا = عَدَا (من الفعل: عَدَا يَعْدُو)، عَدَا (حرف جرّ).
- أو الفعلىة والاسمىة والحرفىة؛ نحو: بل = بَلَّ (فعل)، بَلَّ (مصدر)، بَلَّ (حرف).

لذا، فإن نظام التحليل الصرفى يفترض أن الكلمة التى يعالجها هى فعلٌ واسمٌ وحرفٌ.

وعلى هذا فإن خرج نظام التحليل الصرفى يتضمّن جميع الوجوه الممكنة للكلمة المعالجة. وهى كما ذكرنا أنفاً لا تعدو أن تكون فعلاً أو اسماً أو حرفاً.

مثال: إذا كانت كلمة الدخل هى (سلم)، فإن المحلل الصرفى سيعطى فى خرجه الوجوه الآتية:

- سَلَّمَ (فعل ثلاثى مجرد على وزن فَعَّلَ).
- سَلِمَ (فعل ثلاثى مجرد على وزن فَعِلَ).
- سَلَّمْ (فعل ثلاثى مزيد على وزن فَعَّلَ).
- سَلَّم (مصدر على وزن فَعَّلَ).
- سَلَمَ (اسم على وزن فَعَّلَ).
- سِلْمَ (اسم على وزن فَعَّلَ).
- سَلَمَ (اسم على وزن فَعَّلَ).
- سُلْمَ (اسم على وزن فُعِّلَ).

فإذا كانت الكلمة فعلاً، أعطى المحلل المعلومات الآتية:

- صيغة الفعل (ماضٍ، مضارع، أمر).
- وزنه (فعل، أفعل، فَعَّلَ، تفاعل، استفعل...).
- سابقته (و، ف، س...).
- لاحقته (ضمائر الرفع المتصلة، ضمائر النصب المتصلة).
- جذره.
- الضمير المسند إليه (أنت، هو، نحن...).
- بناءه للمعلوم أو للمجهول.
- تجرّده أو زيادته.
- ضبطه بالشكل التام (بالحركات).

وإذا كانت الكلمة اسماً، أعطى المحلل فى خرجه:

- سابقته (و، أل، ب...).
- لاحقته (ضمير الجر المتصل).
- وزنه (إفعال، فَعِيل، تَفَعَّل...).
- نوع الاسم (مصدر، صفة، ظرف، اسم علم...).
- جذره.
- نوعه من جهة الجمود والاشتقاق (اسم جامد، اسم مشتق "اسم فاعل، اسم مفعول، مبالغة اسم الفاعل...").
- نوعه من جهة التصرف (ممنوع من الصرف، مصروف).

- نوعه من جهة التذكير والتأنيث (مذكر، مؤنث).
- نوعه من جهة الإفراد والتثنية والجمع (مثنى، جمع مذكر سالم، جمع تكسير...).
- نوعه من جهة النسبة.
- نوعه من جهة التصغير.

يبين الشكل الآتي نتيجة تحليل كلمة (فأعمل)، وجزءاً من تحليل كلمة (فكره) في برنامج الخليل الصرفي:

الجهات الراحية والمشاركة في إنجاز نظام الخليل الصرفي:

- إدارة العلوم والبحث العلمي في المنظمة العربية للتربية والثقافة والعلوم.
- مدينة الملك عبد العزيز للعلوم والتقنية، المملكة العربية السعودية.
- جامعة محمد الأول بوجدة، المملكة المغربية.

الخبراء المشاركون في إنجاز نظام الخليل الصرفي:

- د. عز الدين المزروعى، أستاذ بجامعة محمد الأول بوجدة، المغرب.
- د. عبد الحق لخواجة، أستاذ بجامعة محمد الأول بوجدة، المغرب.
- د. عبد الرحيم بودلال، جامعة محمد الأول بوجدة، المغرب.
- أ. محمد ولد عبدالله ولد بيا، جامعة محمد الأول بوجدة، المغرب.
- د. عبد الوافي مزيان، جامعة محمد الأول بوجدة، المغرب.
- د. مصطفى اسهول، جامعة محمد الأول بوجدة، المغرب.
- د. المعتز بالله السعيد، مستشار لغوي وأستاذ بجامعة عين شمس، مصر.
- د. إبراهيم الخراشي، باحث في معهد بحوث الحاسب والإلكترونيات، بمدينة الملك عبد العزيز للعلوم والتقنية، السعودية.
- د. منصور الغامدي، منسق المشروع بمدينة الملك عبد العزيز للعلوم والتقنية، السعودية.
- م. طه زروقي، مستشار في المعلوماتية، الجزائر.
- أ. مروان البواب، عضو مجمع اللغة العربية بدمشق، سوريا.
- أ.د. مروان المحاسني، رئيس مجمع اللغة العربية بدمشق، سوريا.
- أ.د. مكي الحسني، أمين عام مجمع اللغة العربية بدمشق، سوريا.
- د. محمد سعيد دسوقي، أستاذ بالمعهد العالي للعلوم التطبيقية والتكنولوجيا، سوريا.
- د. فاضل سكر، أستاذ بكلية الهندسة المعلوماتية، سوريا.
- د. سمر معطي، أستاذة بكلية الهندسة المعلوماتية، جامعة حلب، سوريا.
- د. حسن السيد، مستشار الجامعة الافتراضية السورية.
- أ. عبد المنعم فريد محمد، مدير شركة مداد لتقنية المعلومات، مصر.
- أ. محمود صابر محمد، مبرمج بشركة مداد لتقنية المعلومات، مصر.
- د. محمد زايد، أستاذ جامعي وخبير بالمنظمة العربية للتربية والثقافة والعلوم.
- أ. د. عبد المجيد حمادو، مدير المعهد العالي للإعلامية والمليمتيديا بجامعة صفاقس، تونس.
- د. جوزيف ديشي، أستاذ في اللسانيات العربية بجامعة ليون (2).
- د. محمد حسون، أستاذ جامعي بجامعة ليون، فرنسا.

- د. رمزي عباس، أستاذ جامعي، فرنسا.
- د. فتحي دبيلي، مدير بحوث بمخبر LLACAN، فرنسا.
- أ. مجدي شاكرا سالم صوالحة، أستاذ في مجال علوم الإعلام، بريطانيا.
- أ. نزار حبش، باحث، الولايات المتحدة الأمريكية.
- أ. تيم باك ولتر، باحث، الولايات المتحدة الأمريكية.
- د. أوتكار سمرز، باحث، تشيكيا.
- د. قاسم سارة، المستشار الإقليمي للإعلام الطبي والصحي بمنظمة الصحة العالمية بمصر.
- د. محمد مراياتي، كبير مستشاري الأمم المتحدة للعلم والتكنولوجيا للتنمية المستدامة.
- د. سلوى السيد حمادة، باحثة بمعهد البحوث الالكترونية بمصر.
- م. نبيل أبو هنية، مركز أنظمة المعلومات، الأردن.
- أ. زياد مرقة، مجموعة طلال أبو غزالة، الأردن.

دليل الاستعمال:

1. تنزيل البرنامج

هذا النظام متاح على موقع **sourceforge.net** وهو أحد المواقع العالمية للبرمجيات الحرة المفتوحة المصدر. وصل عدد تنزيلات النظام إلى **3769** تنزيلاً على مستوى العالم حتى سنة **2014**، منها **2931** تنزيلاً في الدول العربية وتوزعت على النحو التالي:

- عدد تنزيلات النظام بجمهورية مصر العربية: **670**
- عدد تنزيلات النظام بالمملكة العربية السعودية: **584**
- عدد تنزيلات النظام بالمملكة المغربية: **544**
- عدد تنزيلات النظام بالجمهورية الجزائرية الديمقراطية: **311**
- عدد تنزيلات النظام بالجمهورية التونسية: **170**
- عدد تنزيلات النظام بالمملكة الأردنية الهاشمية: **135**
- عدد تنزيلات النظام بجمهورية العراق: **102**
- عدد تنزيلات النظام بالإمارات العربية المتحدة: **79**
- عدد تنزيلات النظام بدولة فلسطين: **65**
- عدد تنزيلات النظام بدولة قطر: **58**
- عدد تنزيلات النظام بدولة الكويت: **58**
- عدد تنزيلات النظام بالجمهورية اليمنية: **35**
- عدد تنزيلات النظام بجمهورية لبنان: **33**
- عدد تنزيلات النظام بسلطنة عمان: **33**
- عدد تنزيلات النظام بدولة ليبيا: **27**
- عدد تنزيلات النظام بمملكة البحرين: **22**
- عدد تنزيلات النظام بموريتانيا: **3**

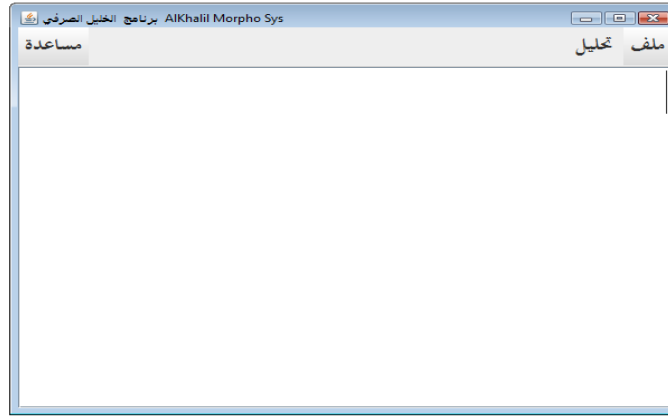
2. التثبيت

وجب تثبيت برنامج آلة جافا الافتراضية ابتداء من النسخة 1.4. كما يمكن استعمال هذا البرنامج مع بيئات متعددة من قبيل Windows و Linux و Max Os و Solaris. يمكن تحميل آلة جافا الافتراضية من موقع Sun Microsystems من خلال

العنوان التالي: <http://java.sun.com>

3. أول استعمال

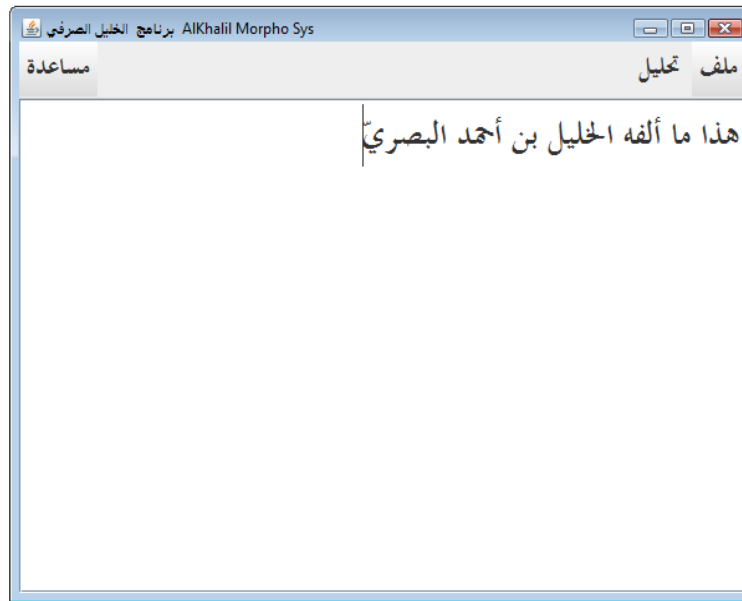
عند تشغيل برنامج الخليل الصرفي، تظهر لوحة إدخال الكلمات الآتية:



يمكن تحليل كلمة أو مجموعة من الكلمات عن طريق كتابتها في مجال النص أو عن طريق تحميلها من ملف خارجي (.txt) مع مراعاة أن يكون ترميز الملف CP-1256.

❖ تحليل كلمة أو مجموعة من الكلمات كتبت في مجال النص:

يكتب المستخدم النص الذي يريد تحليله ثم ينقر على زر "تحليل" في "قائمة الأدوات".



بعد تحليل النص، يحصل المستخدم على النتائج التالية:

نتائج التحليل Analysis Results							حفظ النتائج
الخرج OUTPUT							الدخل INPUT
اللاحق Suffix	الحالة الإعرابية POS Tags	الجزر Root	الوزن Pattern	نوع الكلمة Type	السابق Prefix	الكلمة المشكولة Voweled Word	
#		#	#	اسم إشارة	#	هَذَا	هذا
#	مفرد مذكر منصوب نكرة	هَذَا	فَعَا	مصدر أصلي	#	هَذَا	
#	متعد ولازم (هما) متعد ولازم	هَذَا	فَعَا	فعل ماض مبني للمعلوم	#	هَذَا	
#	متعد (أنتما) متعد ولازم	هَذَا	فَعَا	فعل أمر	#	هَذَا	
#	متعد ولازم (هما) متعد ولازم	هَذَا	فَعَا	فعل ماض مبني للمجهول	#	هَذَا	
#	متعد ولازم (هو) متعد ولازم	هَذَا	هَذَا	فعل ماض مبني للمعلوم	#	هَذَا	
#		#	#	اسم شرط	#	مَا	
#		#	#	اسم موصول	#	مَا	ما

❖ تحليل نص مخزن في ملف:

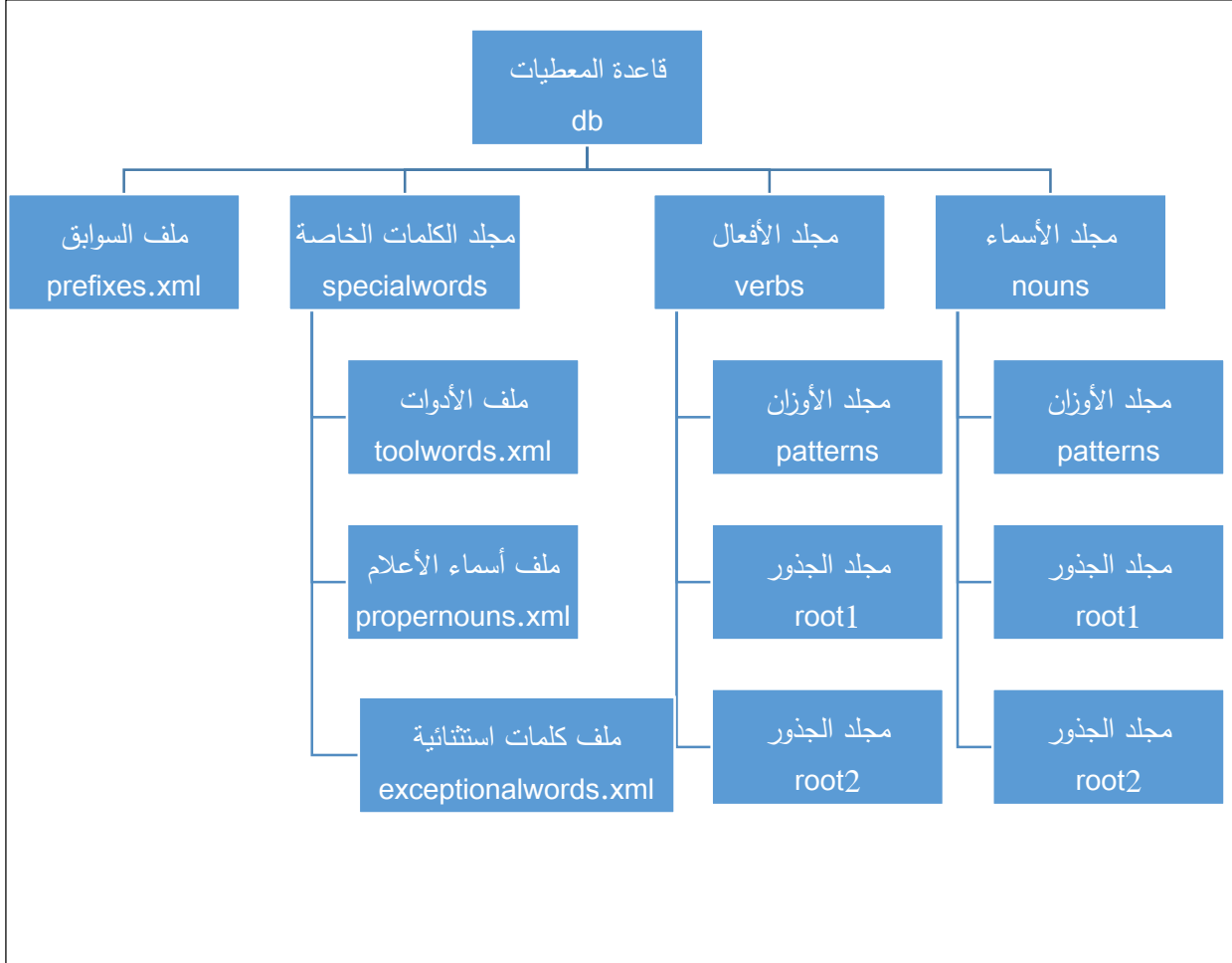
بالنقر في "قائمة الأدوات" على "ملف" ثم على "فتح" ويحدد المستخدم النص الذي يريد تحليله ثم يحمله على البرنامج. ملاحظة: يمكن تحليل جزء معين من النص عن طريق اختياره واتباع الإرشادات السابقة لتحليل النص.

4. حفظ النتائج

النظام يتيح للمستخدم حفظ النتائج في ملف بصيغة HTML وذلك بالنقر على أيقونة "حفظ النتائج".

قواعد المعطيات المستعملة في برنامج الخليل الصرفي:

يعتمد برنامج الخليل الصرفي على قاعدة معطيات تم ترتيبها على شكل مجلدات وملفات بصيغة (XML)، يوضح الشكل التالي البنية الداخلية للمجلد الرئيسي **db** الذي يضم مجموع هذه الملفات:



وفيما يلي مكونات قاعدة بيانات برنامج الخليل الصرفي مفصلة:

❖ ملف السوابق db/prefixes.xml :

يحتوي هذا الملف على لائحة السوابق المستعملة في البرنامج حيث تم إرفاق كل سابق بالمعلومات التالية:

1. السابق دون علامات التشكيل (**unvoweledform**)
2. السابق مشكولا (**voweledform**)
3. التوصيف (**desc**)
4. الصنف (**classe**): حيث تم تصنيف هذه السوابق إلى ثلاثة أصناف:
 - الصنف **N** ويعنى به السوابق التي لا تدخل إلا على الأسماء
 - الصنف **V** ويرمز للسوابق الخاصة بالأفعال

• **الصف C** ويرمز للسوابق المشتركة بين الأسماء والأفعال

كما تم تقسيم كل صنف من هذه الأصناف إلى الأقسام الفرعية التالية:

- **N1**: يضم هذا القسم ال التعريف وتفرعاتها
- **N2**: يضم همزة الاستفهام + ال التعريف وتفرعاتها
- **N3**: يضم لام التوكيد+ ال التعريف وتفرعاتها
- **N4**: يضم حروف الجر وتفرعاتها
- **N5**: يضم حروف الجر + ال التعريف وتفرعاتها
- **V1**: يضم سين المضارعة وتفرعاتها
- **V2**: يضم لام النصب وتفرعاتها
- **V3**: يضم لام الجزم وتفرعاتها
- **C1**: يضم الواو والفاء و " الذي يعني غياب السابق
- **C2**: يضم همزة الاستفهام وتفرعاتها
- **C3**: يضم همزة الاستفهام + ال التعريف وتفرعاتها

أمثلة:

أ.

<prefixe unvoveledform = "ب" voveledform = "بِ" desc = "حرف الجر" classe="N4">

مما يعني أن الباء سابقة تدخل على الأسماء وهي من الصف N4 الخاص بحروف الجر

ب.

<prefixe unvoveledform = "فس" voveledform = "فَسَ" classe="V1">

مما يعني أن الفاء+ سين المضارعة من الصف V1 الخاص بالفعل المضارع المرفوع

ج.

<prefixe unvoveledform = "أ" voveledform = "أُ" classe="C2" >

مما يعني أن همزة الاستفهام سابقة مشتركة بين الأسماء والأفعال.

❖ **اللواحق db/suffixs.xml**

يحتوي هذا الملف على لائحة اللواحق المستعملة في البرنامج وكما هو الحال بالنسبة للسوابق، تم إرفاق كل لاحق بثلاث معلومات

1. اللاحق دون علامات التشكيل (unvoveledform)

2. اللاحق مشكولا (voveledform)

3. التوصيف (desc)

4. الصنف (classe): حيث تم تصنيف اللواحق إلى ثلاثة أصناف:

- الصنف N يرمز للواحق التي لا تدخل إلا على الأسماء
- الصنف V ويرمز للواحق الخاصة بالأفعال
- الصنف C ويرمز للواحق المشتركة بين الأسماء والأفعال

وقد تم تقسيم كل صنف من هذه الأصناف إلى الأقسام الفرعية التالية

- C1: يضم ' الذي يعني غياب اللاحق
- C2: يضم الضمائر البسيطة المسندة إلى المخاطب
- C3: يضم الضمائر البسيطة المسندة إلى الغائب
- N1: ويضم ياء النسبة
- V1: يضم نون الوقاية ياء المتكلم
- V2: يضم نون الوقاية + باقي ما يلحق بها من الضمائر
- V3: يضم الضمائر المركبة من اللواحق من القسم C2 و C3
- V4: يضم واو الجماعة+ ما يلحق بها من الضمائر

أمثلة:

أ.

<suffixe unvoeiledform = "ك" voweiledform = "كَ" desc = "للمخاطب المفرد" classe="C2">

مما يعني أن الكاف لاحقة مشتركة بين الأسماء والأفعال

ب.

<suffixe unvoeiledform = "وهم" voweiledform = "وَهُمْ" desc = "ضمير الغائبين" classe="V4">

مما يعني أن اللاحق واو الجماعة + الضمير تدخل على الأفعال فقط

ج.

<suffixe unvoeiledform = "ي" voweiledform = "ي" classe="N1">

مما يعني أن اللاحق ياء النسبة تدخل على الأسماء فقط.

❖ الأدوات db/specialwords/toolwords.xml

وقد أرفقت كل أداة بالمعلومات التالية:

1. الأداة دون علامات التشكيل (unvoweledform)
 2. الأداة مشكولة (voweledform)
 3. النوع (type):
 4. أقسام السوابق التي تدخل عليها prefixeclass
 5. أقسام اللواحق التي تدخل عليها suffixeclass
 6. أولوية الترتيب وهو رمز يستعمله البرنامج من أجل عرض مخرجات التحليل حسب الترتيب التالي: الأسماء فالأفعال ثم الحروف مع تقديم المعربات على المبنيات.
- فيما يخص الأسماء تم اعتماد الترتيب التالي: العَلم - الاسم الجامد - الضَّمير - اسم الإشارة - الاسم الموصول - اسم الاستفهام - اسم الشرط - الظرف - المصدر الجامد - المصدر الميمي المشتق - اسم المرة / الهيئة - اسم الفاعل - صيغة المبالغة - اسم المفعول - الصفة المشبهة - اسم التفضيل - اسم المكان / الزمان - اسم الآلة
 - فيما يخص الأفعال تم اعتماد الترتيب التالي: ماضي - مضارع - أمر والمبني للمعلوم قبل المبني للمجهول ثم اللازم فالمتعدي واللازم فالمتعدي
 - فيما يخص الحروف تم اعتماد الترتيب التالي: جر - نصب - جزم - عطف - شرط - ناسخ - تحقيق - نداء - استفهام - استثناء

مثال:

```
<toolword unvoweledform="على" voweledform="عَلَى" type="حرف جر" prefixeclass="C1C2C3C4" suffixeclass="C1C2C3N1" priority="301">
```

مما يعني أن كلمة **على** هي حرف جر تدخل عليه السوابق من الصنف C1 و C2 و C3 و C4 واللواحق من الصنف C1 و C2 و C3 و N1 وبالنسبة لترتيبها ضمن لائحة المخرجات فالرمز 3 يعني أنها تعرض في الرتبة الثالثة بعد الأسماء والأفعال والرمز 01 يعني أنها تعرض في الرتبة الأولى ضمن لائحة الحروف شأنها في ذلك شأن باقي حروف الجر.

❖ أسماء الأعلام db/specialwords/propernouns.xml

ويحتوي هذا الملف على 2040 اسمًا علمًا تم ضبطها بعلامات التشكيل كما في المثال التالي:

```
</propernoun unvoweledform="محمد" voweledform="مُحَمَّد"/>
```

❖ الأوزان

وقد تم وضعها في المجلدين db/nouns/patterns الخاص بأوزان الأسماء و db/verbs/patterns الخاص بأوزان الأفعال. ويضم كل منهما مجلدين يحتوي الأول على الأوزان غير المشكولة (db/nouns/patterns/Unvoweled بالنسبة للأسماء و db/verbs/patterns/Unvoweled بالنسبة للأفعال) والثاني على تشكيلات هذه الأوزان (db/nouns/patterns/Voweled بالنسبة للأسماء و db/verbs/patterns/Voweled بالنسبة للأفعال).

▪ الأوزان الغير المشكولة db/nouns/patterns/Unvoweled و db/verbs/patterns/Unvoweled

يحتوي هذان المجلدان على ملفات بصيغة XML تحتوي على أوزان الأسماء والأفعال غير المشكولة. وقد تم تقسيم هذه الأوزان حسب طولها فعلى سبيل المثال يضم الملف UnvoweledNominalPatterns2.xml الأوزان الاسمية غير المشكولة المكونة من حرفين بينما يضم الملف UnvoweledVerbalPatterns7.xml الأوزان الفعلية غير المشكولة المكونة من سبعة حروف.

تم إرفاق كل وزن بالمعلومات التالية

- value: وترمز للوزن السماعي (مثلا فع هو وزن لكلمة مَرّ)
- rules: وترمز لموقع الحروف الأصلية للجذر داخل الوزن مما يمكن من استخلاص الجذر
- ids: ترمز للتشكيلات الممكنة لهذا الوزن

مثال:

```
<pattern value="عل" rules="12ء" ids="14412 14413 14414 14434 14436 14437 14438 14484 14485 14486 14487 14488 14489"/>
```

يوضح هذا المثال أن غياب الفاء في الوزن السماعي عل يمكن أن ينجم عن حذف الواو فنستخدم القاعدة و12 (مثال قف من الجذر وقف) ويمكن أن ينجم عن حذف الهمزة فنستخدم القاعدة ء12 (مثال كل من الجذر عكل). وبالنسبة للتشكيلات المحتملة لهذا الوزن فيعبر عنها عن طريق الرموز 14412 إلى 14489 المستخدمة في ترتيب الأوزان المشكولة في قاعدة معطيات البرنامج.

▪ الأوزان المشكولة db/nouns/patterns/Voweled و db/verbs/patterns/Voweled

يحتوي هذان المجلدان على ملفات بصيغة XML تحتوي على أوزان الأسماء والأفعال المشكولة. وكما هو الشأن بالنسبة لنظيرتها غير المشكولة فقد تم تقسيم هذه الأوزان حسب طولها فعلى سبيل المثال يضم الملف VoweledNominalPatterns3.xml الأوزان الإسمية المشكولة المكونة من ثلاثة أحرف بينما يضم الملف VoweledVerbalPatterns4.xml الأوزان الفعلية المشكولة المكونة من أربعة أحرف.

وتختلف المعلومات المرفقة بكل وزن تبعا لنوع الكلمة حيث قمنا بترميز هذه المعلومات كما هو مبين في الجدولين 1 و 2 المخصصين لتفسير هذه الرموز بالنسبة للاسم والفعل بالترتيب.

بالنسبة للأسماء تم إرفاق الأوزان بالمعلومات التالية

- id: ونعني به الرقم التسلسلي للوزن المشكول في قاعدة المعطيات
- diac: ويمثل الوزن السماعي المضبوط بعلامات التشكيل
- canonic: ويمثل الوزن القياسي المضبوط بعلامات التشكيل

- **type**: ويمثل نوع الاسم التي يشتق منه هذا الوزن
 - **cas**: ويحدد حالة تعريف الاسم من تكثيره
 - **ncg**: ويحدد عدد الاسم وجنسه وحركته الإعرابية
- مثال:

<pattern id="3372" diac="فعل" canonic="فعل" type="صا" cas="نك" ncg="1"/>

مما يعني أن الوزن الاسمي رقم 3372 هو الوزن المشكول فعل (الوزن القياسي في هذه الحالة لا يختلف عن الوزن السماعي). أما نوع الاسم فهو مصدر أصلي نكرة مفرد مذكر مرفوع.

الجدول 1: تفسير الرموز المستعملة في قاعدة معطيات الأوزان الاسمية المشكولة

الرمز	التفسير	الرمز	التفسير
فا	اسم فاعل	1	مفرد مذكر مرفوع
مف	اسم مفعول	2	مفرد مؤنث مرفوع
مفا	مبالغة اسم الفاعل	3	مثنى مذكر مرفوع
آ	اسم آلة	4	مثنى مؤنث مرفوع
زmk	اسم زمان أو مكان	5	جمع مذكر مرفوع
نك	نكرة	6	جمع مؤنث مرفوع
إض	مضاف	7	مفرد مذكر منصوب
فض	اسم تفضيل	8	مفرد مؤنث منصوب
وش	صفة مشبهة	9	مثنى مذكر منصوب
صأ	مصدر أصلي	10	مثنى مؤنث منصوب
صم	مصدر ميمي	11	جمع مذكر منصوب
صر	مصدر مرة	12	جمع مؤنث منصوب
صه	مصدر هيئة	13	مفرد مذكر مجرور
جا	اسم جامد	14	مفرد مؤنث مجرور
صن	مصدر صناعي	15	مثنى مذكر مجرور
نس	نسبة	16	مثنى مؤنث مجرور

وبالنسبة للأفعال تم إرفاق الوزن بالمعلومات التالية

- **id**: ونعني به الرقم التسلسلي للوزن المشكول في قاعدة المعطيات
- **diac**: ويمثل الوزن السماعي المضبوط بعلامات التشكيل
- **canonic**: ويمثل الوزن القياسي المضبوط بعلامات التشكيل
- **type**: ويمثل زمن الفعل التي يشتق منه هذا الوزن
- **aug**: ويمثل حالة الفعل من حيث التجرد والزيادة
- **cas**: ويحدد حركة الفعل المضارع

- **ncg**: ويحدد إسناد الفعل
- **trans**: ويحدد إن كان الفعل لازماً أم متعدياً

مثال:

</> **trans="ك" ncg="2" cas="ج" aug="ضم" type="تَفَعَّل" canonic="تَفَعَّل" diac="تَفَعَّل" pattern id="16877"**

مما يعني أن الوزن الفعلي رقم 16877 هو الوزن المشكول تَفَعَّل (الوزن القياسي في هذه الحالة لا يختلف عن الوزن السماعي). أما نوع الفعل فهو مضارع مبني للمعلوم مجرد مجزوم مسند إلى المتكلمين متعد ولأزم.

الجدول 2: تفسير الرموز المستعملة في قاعدة معطيات الأوزان الفعلية المشكولة

الرمز	التفسير	الرمز	التفسير
م	ماض مبني للمعلوم	1	مسند إلى المتكلم (أنا)
مج	ماض مبني للمجهول	2	مسند إلى المتكلمين (نحن)
ضم	مضارع مبني للمعلوم	3	مسند إلى المخاطب (أنت)
ضمءم	مضارع مؤكد مبني للمعلوم	4	مسند إلى المخاطبة (أنتِ)
ضج	مضارع مبني للمجهول	5	مسند إلى المخاطبين (أنتم)
ضءج	مضارع مؤكد مبني للمجهول	6	مسند إلى المخاطبين (أنتم)
أ	أمر	7	مسند إلى المخاطبات (أنتن)
أء	أمر مؤكد	8	مسند إلى الغائب (هو)
ج	مجزوم	9	مسند إلى الغائبة (هي)
ر	مرفوع	10	مسند إلى الغائبين (هما)
ن	منصوب	11	مسند إلى الغائبتين (هما)
جر	ثلاثي مجرد	12	مسند إلى الغائبين (هم)
زي	ثلاثي مزيد	13	مسند إلى الغائبات (هن)
ك	متعد ولأزم		
م	متعد		
ل	لأزم		

❖ الجذور

تم وضع الجذور المستعملة في البرنامج مرفقة بأرقامها التسلسلية في ملفين: يحتوي الملف الأول db/Allroots1.txt على قاعدة موسعة من الجذور تضم 7502 جذراً ويحتوي الملف الثاني db/Allroots2.txt على قاعدة مختصرة تحتوي على 2900 جذر. كما تم إرفاق هذه الجذور بأوزان مشتقاتها وتم وضعها في المجلدات الأربعة التالية:

- المجلدان db/nouns/roots1 و db/verbs/roots1 الخاصان بالقاعدة الموسعة للجذور الاسمية والفعلية والتي تضم. تشير إلى أن جذور هذين المجلدين.

• المجلدان db/nouns/roots2 و db/verbs/roots2 الخاصان بالقاعدة المختصرة للجذور الاسمية والفعلية شائعة الاستعمال والتي تضم 2900 جذرا.

تحتوي هذه المجلدات على ملفات بصيغة XML مخصص كل واحد منها للجذور التي تشترك في الحرف الأول. وقد تم إرفاق كل جذر بالمعلومات التالية:

- **val**: وترمز للجذر
- **vect**: ويضم الأرقام التسلسلية لجميع الأوزان المشكولة لمشتقات هذا الجذر

مثال:

```
<root val="كتب" vect="10608 10610 10616 10618 10620 10622 10624 10673 10674 10675 10676 10677 10678 10679  
10680 10681 10682 10683 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 .....  
8666 8667 8668/>
```

يبين هذا المثال جزءا من الأرقام التسلسلية للأوزان المشكولة لمشتقات الجذر كتب