

برنامج الخليل المصرفي

الإصدار الأول

2010

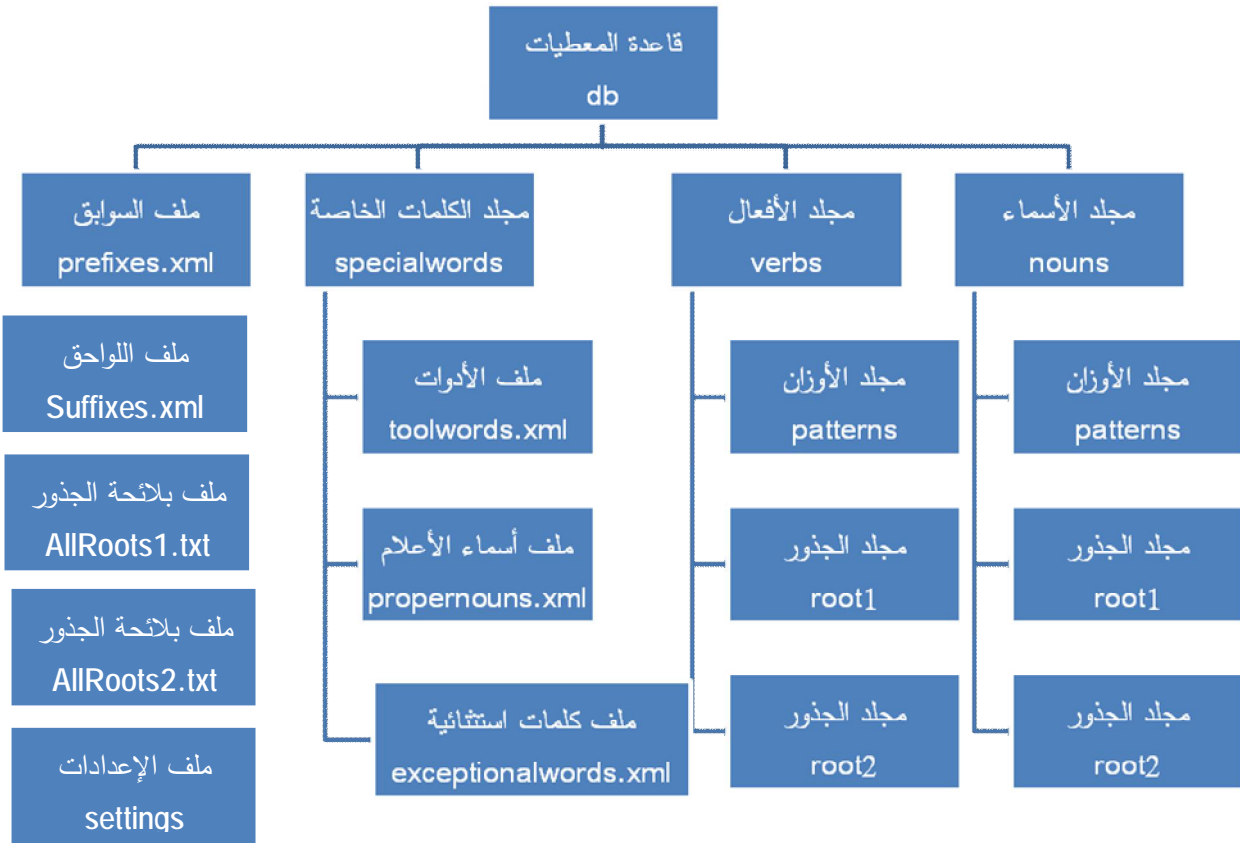
ورقة حول قواعد المعطيات المستعملة

في برنامج الخليل الصرفي

تقديم

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 نعرض لمكونات هذه القاعدة بشيء من التفصيل .

1. ملف السوابق db/prefixes.xml :

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

أ- السابق دون علامات التشكيل (unvoweledform)

ب- السابق مشكولا (voweledform)

ج-التوصيف (desc)

د- الصنف (classe): حيث تم تصنيف هذه السوابق إلى ثلاثة أصناف:

• الصنف N ونعني به السوابق التي لا تدخل إلا على الأسماء

• الصنف V ويرمز للسوابق الخاصة بالأفعال

• الصنف C ويرمز للسوابق المشتركة بين الأسماء والأفعال

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

- N1: ويضم 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 unvoweledform = "ب" voweledform = "بِ" desc = "حرف الجرف"
classe="N4">

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

ب.

<prefixe unvoweledform = "فس" voweledform = "فَسَ" classe="V1">

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

ج.

<prefixe unvoweledform = "أ" voweledform = "أَ" classe="C2" >

مما يعني أن همزة الاستفهام سابقة مشتركة بين الأسماء والأفعال

2. اللواحق db/suffixs.xml

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

أ- اللاحق دون علامات التشكيل (unvoweledform)

ب- اللاحق مشكولا (voweledform)

ج-التوصيف (desc)

د- الصنف (classe): حيث تم تصنيف اللواحق إلى ثلاثة أصناف:

• الصنف N ونعني به اللواحق 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 unvoweledform = "ك" voweledform = "كَ" desc = "للمخاطب المفرد" classe="C2">

مما يعني أن الكاف لاحقة مشتركة بين الأسماء والأفعال
ب.

<suffixe unvoweledform = "وهم" voweledform = "وَهُمْ" desc = "ضمير الغائبين" classe="V4">

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

<suffixe unvoweledform = "ي" voweledform = "ي" classe="N1">

مما يعني أن اللاحق ياء النسبة تدخل على الأسماء فقط

3. الأدوات db/specialwords/toolwords.xml

وقد أرفقت كل أداة بالمعلومات التالية:

أ - الأداة دون علامات التشكيل (unvoweledform)

ب - الأداة مشكولة (voweledform)

ج - النوع (type):

د - أقسام السوابق التي تدخل عليها prefixclass

هـ - أقسام اللواحق التي تدخل عليها suffixclass

و - 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: العَم - الاسم الجامد - الضمير - اسم الإشارة - الاسم الموصول - اسم الاستفهام - اسم الشرط - الظرف - المصدر الجامد - المصدر الميمي المشتق - اسم المزة / 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

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

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

4. أسماء الأعلام db/specialwords/proper nouns.xml

ويحتوي هذا الملف على 2040 اسما علما تم ضبطها بعلامات التشكيل كما في المثال التالي:

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

5. 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/verbs/patterns/Voweled و db/nouns/patterns/Voweled بالنسبة للأفعال)

5.1 الأوزان غير 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 مَرَّ)
- rules: وترمز لموقع الحروف الأصلية للجذر داخل الوزن مما يمكن من استخلاص الجذر
- ids: ترمز للتشكيلات الممكنة لهذا الوزن

مثال:

```
<pattern value="عل" rules="12 و 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 12ء (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.

5.2. 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 تنكيه
- **ncg**: ويحدد عدد الاسم وجنسه وحركته الإعرابية

مثال:

```
<pattern id="3372" diac="فَعْلٌ" canonic="فَعْلٌ" type="صأ" cas="نك" ncg="1"/>
```

مما يعني أن الوزن الاسمي رقم 3372 هو الوزن المشكول فَعْلٌ (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
زمك	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
		15	مثنى مذكر مجرور
		16	مثنى مؤنث مجرور
		17	جمع مذكر مجرور
		18	جمع 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 لازما أم متعديا

مثال:

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

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)
ضءم	مضارع مؤكد مبني للمعلوم	4	مسند إلى المخاطبة (أنت)
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 الغائبين (هما)
جر	ثلاثي مجرد	12	مسند إلى الغائبين (هم)
زي	ثلاثي مزيد	13	مسند إلى الغائبات (هن)
ك	متعد ولازم		
م	متعد		
ل	لازم		

6. 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 جذرا 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="نصر" 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 نصر