

## البحث ما وراء الكلمات

الدكتور مهندس/ حسام الدين محجوب، رئيس شركة الخوارزمي للبرمجيات  
الأستاذ/ ياسر محمد شعبان، عضو مجلس الإدارة ورئيس قسم الشؤون اللغوية

### ملخص

مع التطور الهائل في الأونة الأخيرة في مجال تطبيقات الحاسوب، ومع الأهمية الراسخة للغة العربية في كافة مجالات الحياة ومنها الجانب التقني، أصبح هناك كم كبير من المحتوى العربي، سواء على المواقع أم على الشبكات الخاصة بالمنظمات والشركات. وللاستفادة من المعلومات الموجودة في هذا المحتوى، تجلّت ضرورة الملحة للبحث داخله.

إن نظم البحث التقليدية والمتبعة في اللغات الأجنبية قاصر إلى حد كبير فيما يخص اللغة العربية نتيجة لقواعدها الصرفية، مما لا يعطي المستخدم - أو الباحث العربي - النتائج المرجوة؛ فثمة مشكلتان: الأولى هي عدم شمول نتائج البحث، والثانية ظهور نتائج ليس لها صلة بكلمات البحث المطلوبة. بالإضافة إلى ذلك فإن خوارزميات ترتيب نتائج البحث في المحركات الأجنبية تعطي في الصفحات الأولى نتائج علاقتها ضعيفة بمعاني كلمات البحث المطلوبة.

ولذا فقد قمنا على مدى العشرين عامًا الماضية بتطوير محرك بحث يراعي طبيعة وخصائص اللغة العربية مما يعود على الباحث العربي بنتائج شاملة ودقيقة، من ناحية، وتطوير خوارزمية تسمح بترتيب نتائج البحث حسب المعنى المطلوب من ناحية أخرى.

فعلى الصعيد الأول طورنا محللاً صرفياً يراعي القواعد الصرفية للغة العربية، بما يحلّ المشكلة الأولى فيحقق الشمول في نتائج البحث. ويستعين المحلل الصرفي بقاموس إلكترونيّ به صفات صرفية ومرادفات ومجالات دلالية للكلمات وهو مبنيّ على "معجم اللغة العربية المعاصرة" للأستاذ الدكتور/ أحمد مختار عمر (رحمه الله)، أستاذ علم وصناعة المعاجم العربية في العصر الحديث.

ولكن القواعد الصرفية وحدها لا تعالج المشكلة الثانية لأن الأشكال الصرفية قد تعطي نتائج بحث بها كلمات ليس لها علاقة دلالية (تتعلق بالمعنى) بكلمات البحث المدخلة، بل لها علاقة صرفية (تتعلق بالشكل) فقط؛ والباحث العربي يبحث عن نتائج لها نفس "المعنى" لكلمات البحث.

ولذلك أثرينا محرك البحث بخوارزميات تفرّق بين النتائج حسب المعنى (أو المجال الدلالي) على مستوى المفردات، مع مراعاة ترتيب نتائج البحث بحيث تظهر العبارات الأقرب لمعاني كلمات البحث أولاً، لتظهر النتائج التي لها علاقة بالكلمات موضع البحث في الصفحات الأولى للنتائج. وكل هذه الابتكارات تعطي دقة أكثر في استرجاع نتائج البحث، فيعالج المشكلة الثانية.

يعتمد البحث الافتراضي على البحث بالمشتقات، أما خيارات البحث فتتضمن البحث بالأشكال الصرفية وبالمرادفات، بالإضافة إلى البحث بالكلمة كما هي - إذا رغب المستخدم في ذلك - والبحث بالعلاقة المنطقية "أو". كما أن نظام إدخال كلمات البحث يطابق تماماً محركات البحث الأجنبية.

## المقدمة

مع التضخم الذي يشهده المحتوى العربي على الشبكة العنكبوتية ومع المبادرات العديدة التي تدعو إلى المزيد من الإثراء للمحتوى العربي، يتحتم وجود أداة للبحث في هذا الكم الضخم بطريقة تناسب الباحث العربي بحيث يتم استرجاع معلومات لها فائدة ومعنى للمستخدم العربي.

لذا يجب أن تراعي أداة البحث صفات كلمات اللغة العربية التي لا غنى عنها حتى يشعر الباحث العربي بأن المعلومات المسترجعة لها علاقة قوية بما يبحث عنه؛ وهذا ما لا يجده المستخدم عندما يبحث في المحتوى العربي باستخدام محركات البحث التقليدية التي صُممت للغات الأجنبية في الأساس، سواءً كانت هذه المحركات موجودة داخل المواقع العربية المختلفة، أم كانت داخل الشركات والمؤسسات العربية، أم كانت محركات بحث تضم جميع المواقع على الشبكة العنكبوتية.

فبعض هذه المحركات، بأنواعها المختلفة، تبحث بالأشكال الإملائية فقط لكلمات البحث وبعضها الآخر يبحث بمشتقات الكلمة العربية فقط؛ وجدير بالذكر هنا أن بعض المواقع المتخصصة مثل مواقع القرآن الكريم والنصوص الدينية تتوفر فيها محركات بحث تراعي الأشكال الصرفية للكلمات العربية على أساس الجذر بالإضافة إلى المشتقات، ولكن هذه المواقع قليلة ولا تكفل السواد الأعظم من المواقع العربية.

وقد شهدنا على مدى السنوات الماضية محاولات عديدة لطرح محركات بحث "عربية" على الشبكة العنكبوتية بها إمكانية البحث بالأشكال الصرفية للكلمات ولكنها لم تحظ بقدر من الانتشار، بل وتم إغلاق الكثير منها. وقد احتكنا وتعاملنا مع بعض هذه المواقع - بطلب ممن يقومون على إدارة هذه المواقع - بغرض تركيب تقنية البحث اللغوي المبتكر الذي قمنا بتطويره داخل هذه المحركات، ولكن هذه المحاولات لم تتم لأسباب، بعضها تسويقية وبعضها فنية؛ وبذلك تعرّفنا على بعض أسباب عدم انتشار هذه المحركات.

ومن أهم الأسباب التقنية التي حالت دون انتشار هذه المحركات هو أن هذه المحركات تستخدم، في بنيتها الأساسية، بعض العناصر مثل محركات أساسية تُستخدم للغات الأجنبية (عن طريق الشراء أو الاستئجار) ثم تقوم بتطوير محلل صرفي وتركيبه على المحرك الأساسي؛ حتى أن البعض كان يستأجر محللا صرفيا عربيا قامت جهات أجنبية بتطويره. وقد نتج عن هذا أمران هامان أدّيا إلى عدم انتشار هذه المحركات:

1. لم تكن هناك مزايا معيّنة تميزها عن محركات البحث التقليدية الأجنبية ما عدا إمكانية البحث بجميع الأشكال الصرفية؛ ونرى أنه لا بد وأن تتوفر ميزة يمكن أن نسميها بـ "روح اللغة" والتي تجعل الباحث العربي يشعر بأن نتائج البحث شاملة وأن لها علاقة بما "يعنيه". وسوف نرى فيما بعد كيف أمكننا إدخال هذه الخاصية في محرك البحث الذي طورناه وأن البحث بالأشكال الصرفية فقط يُقلّل من دقة نتائج البحث.

2. كان يصعب - بل ويستحيل - تعديل أو تطوير هذه المحركات لاعتمادها على برامج أخرى مغلقة المصدر.

أما ما يخصّ المحرك الذي نعرضه فقد قمنا بتطويره بالكامل ونملك مصدره، ما يمكننا من تعديله بإضافة صفات لغوية جديدة، أو بتطويره كي يعمل على أجهزة عديدة مثل الهواتف المحمولة والحاسوبات اللوحية إلخ. وبالفعل تمكّنّا - بناءً على طلب من جهة من الجهات التي تعمل في مجال التعليم الإلكتروني في المملكة العربية السعودية - من إضافة إمكانية البحث بأوزان كلمات البحث.

### مشكلة البحث بالكلمات العربية باستخدام محركات البحث التقليدية

يجب أن يستوفي أي محرك بحث عددا من الشروط الأساسية لضمان جودة نتائج البحث:

- الشمول: وهذا يعني أن المحرك يسترجع جميع النتائج التي لها علاقة بكلمات البحث.
- الدقة: وهذا يعني أن المحرك لا يسترجع نتائج ليس لها علاقة بكلمات البحث.
- الترتيب: يجب أن تُرتب نتائج البحث بحيث تظهر العبارات الأقرب "علاقةً" لكلمات البحث أولاً.

وهذه الشروط يجب أن تنطبق على محركات البحث الموجودة داخل المواقع أو داخل شبكات الشركات والمؤسسات أو على محركات البحث التي تغطي الشبكة العنكبوتية.

إن محركات البحث التقليدية للغات الأجنبية (على الشبكة العنكبوتية، مثلاً) تستوفي هذه الشروط إلى حد كبير بالنسبة للغة الإنجليزية وبعض اللغات الأوروبية، ولكنها لا تستوفيها بالقدر المطلوب بالنسبة لبعض اللغات الأخرى، ما ألجأ بعض هذه الدول لتطوير محركات خاصة تراعي خصائص لغة البلد التي تنتمي لها، مثل محرك "بايدو" للغة الصينية ومحرك "يندكس" الروسي ومحرك "نافر" للغة الكورية وغيرها.

وهذا القصور في استيفاء الشروط المذكورة ينطبق بقدر كبير أيضاً على البحث باللغة العربية، مما يُحتم وجود تقنية بحث عربيّة لمعالجة هذه المشكلة. فالتقنيات "وليدة" بينها ولا يمكن "نقل" التكنولوجيا من بيئة إلى أخرى وإلا حدث نوع من الخلل. لذا فإن استخدام محركات بحث طوّرت للغة أجنبية معيّنة لا يمكن أن تستوفي الشروط المطلوبة إذا ما طبّقت للبحث في المحتوى العربي.

### الخطوة الأولى: استيفاء شرط الشمول - استخدام القواعد الصرفية

لاستيفاء شرط الشمول طوّرنا محلاً صرفياً للكلمات العربية وتم دمجها في محرك البحث بحيث يمكن إيجاد جميع النتائج التي بها كلمات تنتمي لنفس الجذر المعجمي، مع اختلاف صيغها الصرفية. فمثلاً، إذا كانت كلمة البحث "اجتمع" فيتم العثور على الكلمات الآتية (إذا وُجدت في النص): الاجتماع - اجتماعات - الاجتماعات - يجتمعوا - اجتمعوا - مجتمعون... إلخ.

وعلى سبيل المثال، إذا أدخل المستخدم كلمات البحث الآتية: "اجتماع اللجنة العامة"، قد تُسترجع النتائج الآتية:

- ".....وقد اجتمعت أمس الأول اللجنة العمومية لمنظمة الوحدة الإفريقية....."
- ".....إن اللجان الفرعية والعامّة سوف تقوم بالاجتماع في الشهر القادم....."

إن هذا الأسلوب في البحث أكثر ملاءمة لطبيعة اللغة العربية، وهو يستوفي شرط الشمول الذي تم نقاشه في الفقرة السابقة.

ولكن، ماذا عن الدقة في إيجاد نتائج البحث؟

للإجابة على هذا السؤال، سوف نقوم بعرض الصيغ الصرفية المختلفة المندرجة تحت جذر من الجذور؛ ولنأخذ مثلاً الجذر (ج م ع) ولنستعرض بعضاً من الصيغ الصرفية التابعة له: جَمَعَ - أُجْمِعَ - اجْتَمَعَ - إجماع - جماعة - اجتماع - اجتماعية - جمعية... إلخ.

من هنا يتضح أنه إذا تم إدخال كلمة البحث "اجتمع" فيمكن أن نجد نتائج البحث متضمنة الآتي:

- " ... وبعد الاجتماعات القادمة للمجلس .... "
- " ... وكان أعضاء المجلس قد وافقوا بالإجماع على انتخاب ... "
- " ... وقررت وزارة الشؤون الاجتماعية منح الجمعية مهلة .... "

إن شرط الدقة لا يزال غير مستوفى لدرجة كبيرة، حتى مع استخدام القواعد الصرفية للكلمات؛ إذن لابد من إيجاد حل لتحسين دقة البحث وتقليل عدد الكلمات التي ليست لها علاقة دلالية (تتعلق بالمعنى) بكلمة البحث.

**الخطوة الثانية: استيفاء شرط الدقة - استخدام خوارزميات للتعرف الدلالي على الكلمات**

يجب أولاً أن نفكر فيما نعنيه حينما نتكلم عن النتائج المطلوبة. إن المستخدم عندما يبحث، فإنه يضع في ذهنه معنى معيناً، ولذلك فمن الواجب تحديد هذا المعنى، عندما يحدث نوع من الملاسة بين الكلمات.

فمثلاً، في جملة البحث "المتحف الحديث" لیس في كلمة "الحديث"؛ إذ إنها قد تعني <الحدثاء> أو <الكلام>، عندئذ يجب تحديد المعنى المطلوب من قِبَل المستخدم.

فإذا تم اختيار المعنى <الحدثاء>، يتم إيجاد الجملة الآتية في النص مثلاً: " ... **المتاحف الحديثة** في العالم العربي .... " وغيرها من الجمل التي تحتوي على إحدى الصيغ الصرفية لكلمة "الحديث" والتي تنتمي فقط للمعنى <الحدثاء>.

أما إذا تم اختيار المعنى <الكلام>، فلن يتم العثور على النتيجة السابقة وإنما تظهر نتائج مثل: " ... وفي **المتحف** قال المتحدثون الرسميون من مجلس الإدارة ..... "، بالإضافة إلى الجمل التي تحتوي على صيغ "الحديث" التي تنتمي للمعنى <الكلام> فقط.

ونرى من هذه المعالجة أنه يتم تضييق نطاق النتائج لما هو أقرب للمعنى المطلوب من المستخدم، وبذلك يتم الحصول على المعلومات المطلوبة بدقة أكثر من استخدام القواعد الصرفية للغة بمفردها، حيث يقوم النظام المطور بعرض المعاني المختلفة لكل كلمة بحث يتم إدخالها وعرضها على المستخدم، ليقوم باختيار المعنى الذي يقصده. ولتحقيق هذه الخطوة، تم تطوير أداة تقوم بالتعرف الدلالي على الكلمات وتتفاعل هذه الأداة مع المحلل الصرفي.

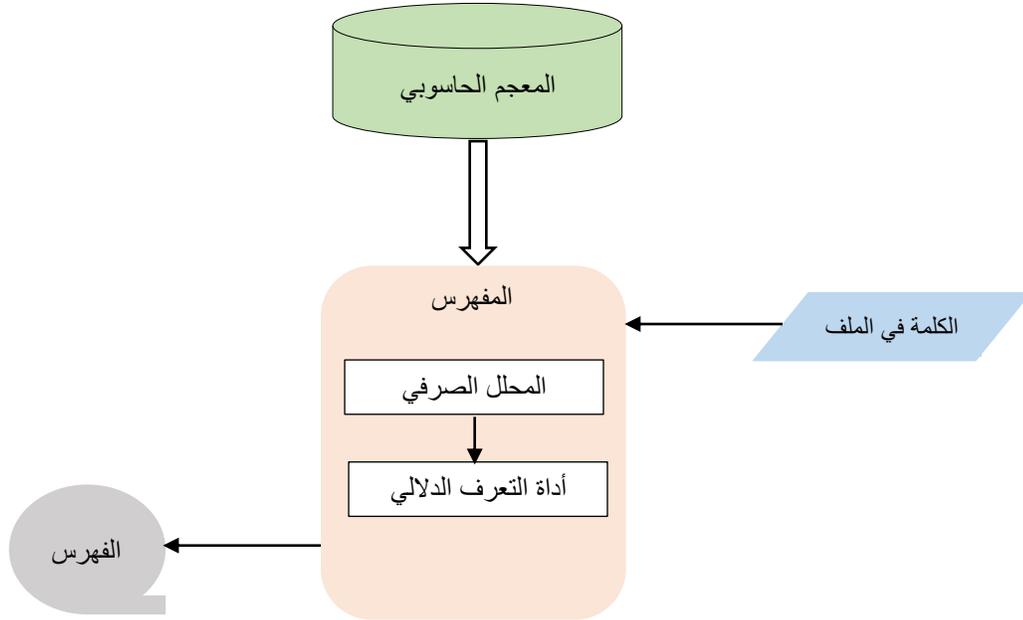
بالإضافة إلى ذلك، تم تطوير خوارزميات تقوم بترتيب نتائج البحث بحيث يتم عرض العبارات الأقرب لمعاني كلمات البحث أولاً وبذلك تُعرض النتائج التي لها علاقة صرفية ودلالية في نفس الوقت، بالكلمات موضع البحث، في الصفحات الأولى للنتائج. وبطبيعة الحال تتفاعل هذه الخوارزميات مع المحلل الصرفي وأداة التعرف الدلالي للكلمات.

### مكونات النظام المتكامل

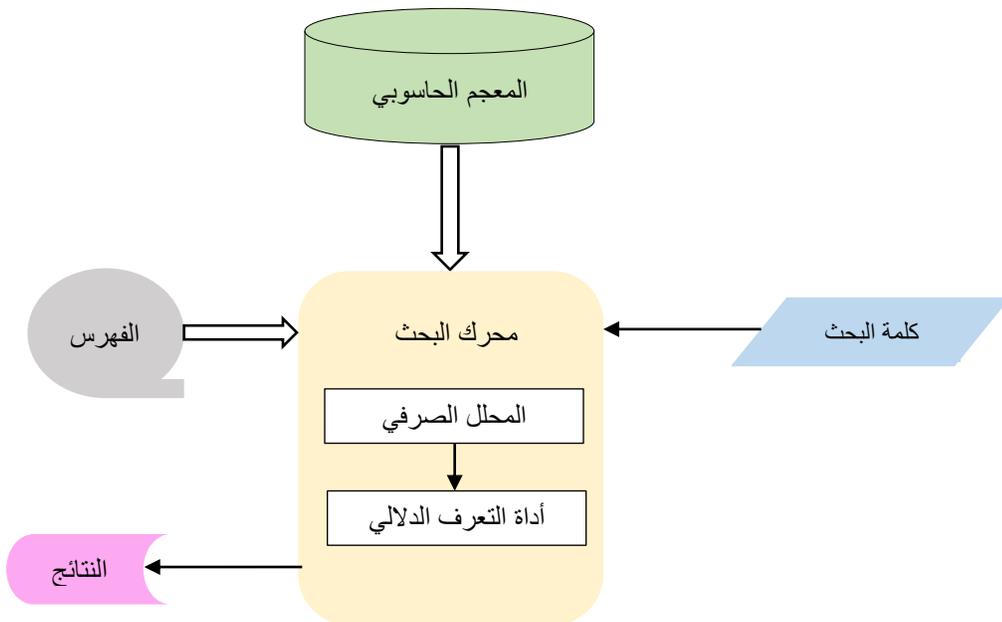
يبين الشكل رقم (1) كيف ينقسم النظام البحثي المتكامل الذي تم تطويره إلى عناصر أو مكونات، ومنه نرى وصفاً لوظيفة كلٍّ منها والعناصر الأخرى التي يتفاعل معها. ويبين الشكل رقم (2) كيفية تفاعل كلٍّ من المكونات مع الآخر أثناء عملية الفهرسة والشكل رقم (3) أثناء عملية البحث.

العناصر التي يتفاعل معها	الوظيفة	العنصر/المكون
- جميع المكونات الأخرى.	قاموس إلكتروني شامل للغة العربية المعاصرة، يحتوي على مداخل الكلمات العربية وصفاتها الصرفية بالإضافة إلى بعض الصفات الإعرابية والدلالية.	المعجم الحاسوبي
- المعجم الحاسوبي. - المحلل الصرفي. - أداة التعرف الدلالي.	يقوم المفهرس بفهرسة الملفات قيد البحث حتى يتسنى البحث فيها فيما بعد بسرعة عالية؛ والفهارس الناتجة عن عملية الفهرسة شبيهة بفهارس الكتب والمعاجم إلخ.	المفهرس
- المعجم الحاسوبي. - الملفات (أثناء الفهرسة). - الكلمات المُدخلة (أثناء البحث).	التحليل الصرفي للكلمات العربية أثناء فهرسة الملفات أو التحليل الصرفي لكلمات البحث المُدخلة.	المحلل الصرفي
- المعجم الحاسوبي. - المحلل الصرفي.	التعرف على المجالات الدلالية (المعاني) للكلمات أثناء فهرسة الملفات وللحصول على الكلمات المُدخلة أثناء البحث.	أداة التعرف الدلالي على الكلمة
- المعجم الحاسوبي. - المحلل الصرفي. - أداة التعرف الدلالي.	البحث في الملفات التي تمت فهرستها.	محرك البحث

الشكل رقم (1): جدول مكونات النظام المتكامل



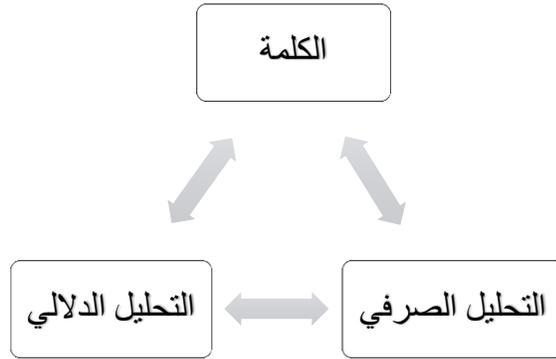
الشكل رقم (2): تفاعل المكونات أثناء عملية الفهرسة



الشكل رقم (3): تفاعل المكونات أثناء عملية البحث

## التعريف بالمعجم الحاسوبي

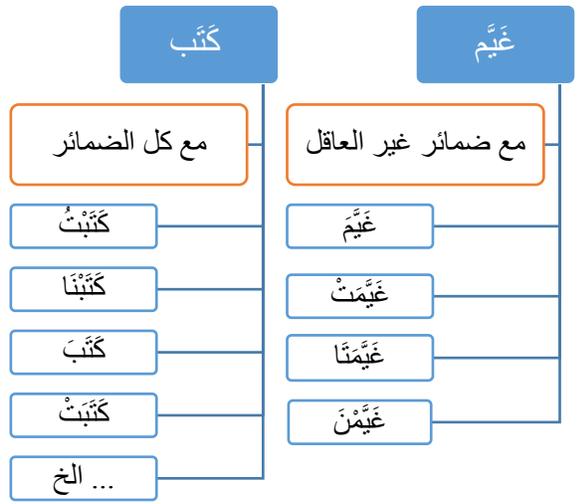
يمثل المعجم الحاسوبي التفاعل بين علم اللسانيات وعلم المعلوماتية، فهو في الأساس يقوم على معجم اللغة العربية المعاصرة الذي قام بإعداده الأستاذ الدكتور أحمد مختار عمر-رحمه الله-، فالمعجم يحتوي على مداخل تستوعب مفردات العربية المعاصرة، ومن الجانب الآخر تقوم المعالجة الإلكترونية لهذه المداخل بإنتاج كم هائل من الأشكال الصرفية للكلمات العربية وربطها بالمجال الدلالي، فالمعجم يعمل وفق المستويين؛ الصرفي والدلالي.



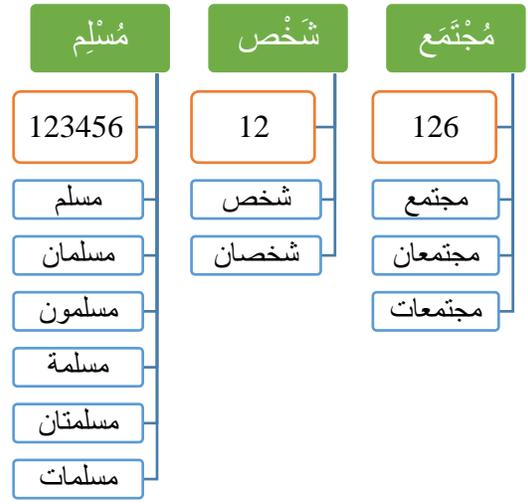
## المدخل/الجذر

وهو المحور الذي تدور حوله عملية توليد الأشكال الصرفية للكلمة، ويمثل الكلمة غير المصرفة (في شكلها المجرد)، وتُحدّد للمدخل المعلومات الآتية:

1. **جذر الكلمة:** وهو الجذر اللغوي للمدخل، ومن ثمّ هو الجذر لجميع مشتقاته، ويجمع الجذر أكثر من مدخل معجمي.
2. **نوع الكلمة:** ومنه يتم التعرف على بعض البيانات حول الكلمة؛ اسم، فعل، حرف، كما يتم التعرف على تصنيف الاسم من حيث الجمود والاشتقاق، ونوع المشتق، والفعل من حيث الماضي والمضارع والأمر.
3. **الصفة الصرفية الأولى:** أداة مهمة لتحديد المشتقات التي تتولد آلياً من الوحدات المعجمية؛ فيما يخص الاسم: للتعرف على بعض المعلومات عن الاسم كالعدد والنوع، وفيما يخص الفعل: لتصريف الفعل حسب ضمائر الفاعل، مع مراعاة الاستعمال اللغوي:



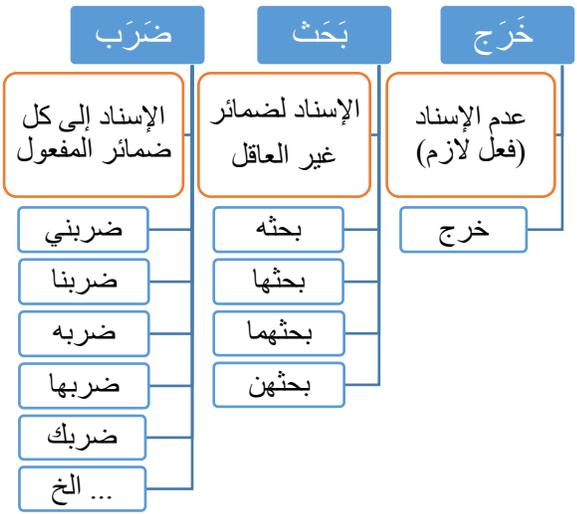
الشكل (5): نظام تصريف الفعل مع ضمائر الفاعل



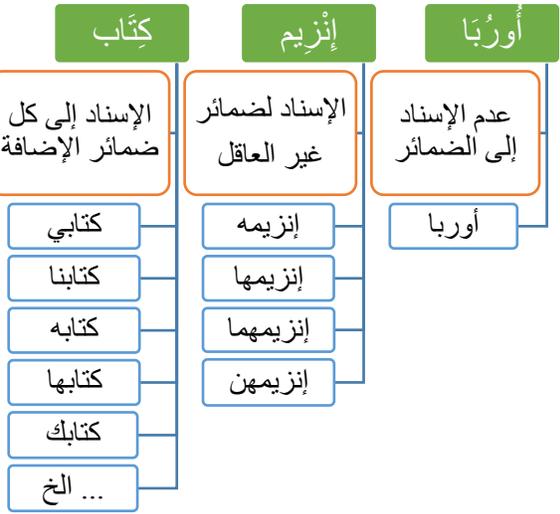
الشكل (4): نظام تصريف الاسم حسب النوع والعدد

#### 4. الصفة الصرفية الثانية:

لأننا نراعي الشمول؛ فكان لابد من مراعاة إسناد الأسماء إلى ضمائر الإضافة، وكذا إسناد الأفعال إلى ضمائر المفعول، وهنا يتم التحكم في صياغة وتوليد الكلمات المسندة إلى هذه الضمائر، مع الأخذ في الاعتبار أن عملية الإلحاق هذه لن تكون عشوائية وإنما تأتي تبعاً للاستعمال اللغوي، فليس كل الأسماء سواءً في عملية الإسناد للضمائر، وليس الفعل اللازم (الذي لا يسمح بإسناد ضمائر المفعول) كالمفعول (الذي يتوقع أن تلحق به ضمائر المفعول)، انظر الشكلين التاليين:



الشكل (7): نظام تصريف الفعل مع ضمائر المفعول



الشكل (6): نظام إسناد الاسم إلى ضمائر الإضافة

## 5. تصنيف الوحدة المعجمية:

مراعاةً للأشكال البنوية المختلفة للكلمات العربية وتعدد أوزانها، تم تصنيف جميع مداخل المعجم الحاسوبي إلى مجموعات من هذه التصنيفات، حيث تتم معالجة كل مجموعة وفق نظام محدد، ولنأخذ الأمثلة التالية، لتوضيح الفرق بين تصنيف أفعال مختلفة البنية مع ضمير المتكلم (أنا):

الفعل مجرداً	الفعل مُسنّداً	الإجراء
كَتَبَ	كَتَبْتُ	إضافة اللاحقة (ت) دون تغيير في الفعل.
قَالَ	قُلْتُ	إضافة اللاحقة (ت) وحذف عين الفعل.
مَدَّ	مَدَدْتُ	إضافة اللاحقة (ت) وفك التضعيف.
قَضَى	قَضَيْتُ	إضافة اللاحقة (ت) وقلب حرف العلة ياءً.
عَزَا	عَزَوْتُ	إضافة اللاحقة (ت) وقلب حرف العلة واوًا.

الشكل السابق يمثل حالة واحدة للإسناد لبعض أنواع الفعل الماضي الثلاثي (فعل)، فإذا أخذنا في الاعتبار مختلف تصنيفات الأفعال بمختلف أوزانها وإسنادها لكل الضمائر (وكذلك إسناد مختلف الأسماء)، يتطلب الأمر معالجة خوارزمية واسعة تستند على جداول مُعدّة صرفياً تستوعب كافة تصنيفات وأوزان هذه المجموعات المحددة في المعجم.

وتعمل هذه الخوارزمية على هذه الجداول الصرفية جنباً إلى جنب مع البيانات المدخلة في المعجم الحاسوبي، وكذلك مع مجموعات مُعدّة ومصنفة من السوابق التي تخص الأفعال والأسماء، والتوافق بين السوابق والضمائر الملحقة، ومن خلال هذه العملية تنتج الأشكال الصرفية الصحيحة للكلمة العربية، والشكل رقم (8) يمثل أحد الجداول الصرفية المُعدّة للفعل الماضي:

حيث يمثل العمود الثالث من الجدول (القلب) شكل ساق الفعل عند الإسناد إلى الضمائر المختلفة، ويمثل العمود الرابع (لواحق 0) لواحق الفعل، عند إسناده إلى ضمير الفاعل فقط، وتمثل بقية الأعمدة ابتداءً من العمود الخامس (لواحق >0، 1، 2، ... 11) لواحق الفعل، عند إسناده إلى ضمير الفاعل وضمير المفعول.

التصنيف: ف1_01					1	كَتَبَ يَكْتُبُ - أَكْتُبُ					x-x-x	الفعل السالم			
هَنْ	هَم	هَمَا	هَا	ه	كُنَّ	كُم	كَمَا	كِ	كَا	كِي	لواحق		القلب	الضمير	
11	10	9	8	7	6	5	4	3	2	1	>0	0			
0001	0001	0001	0001	0001	0001	0001	0001	0001	0010	0100	←	1000			
1	1	1	1	1	1	1	1	1	1	1	ثَا	ثَا	فَعَلْ	أنا	1
1	1	1	1	1	1	1	1	1	1		ثَا	ثَا	فَعَلْ	نحن	2
1	1	1	1	1					1	1	ثَا	ثَا	فَعَلْ	أنت	3
1	1	1	1	1					1	1	ثَا	ثَا	فَعَلْ	أنت	4
1	1	1	1	1					1	1	ثَمَا	ثَمَا	فَعَلْ	أنتما	5
1	1	1	1	1					1	1	ثُم	ثُم	فَعَلْ	أنتم	6
1	1	1	1	1					1	1	ثُنَّ	ثُنَّ	فَعَلْ	أنتن	7
1	1	1	1	1	1	1	1	1	1	1	.	×	فَعَلْ	هو	8
1	1	1	1	1	1	1	1	1	1	1	ثَا	ثَا	فَعَلْ	هي	9
1	1	1	1	1	1	1	1	1	1	1	ا	ا	فَعَلْ	هما (مذكر)	10
1	1	1	1	1	1	1	1	1	1	1	ثَا	ثَا	فَعَلْ	هما (مؤنث)	11
1	1	1	1	1	1	1	1	1	1	1	وا	وا	فَعَلْ	هم	12
1	1	1	1	1	1	1	1	1	1	1	نَا	نَا	فَعَلْ	هن	13

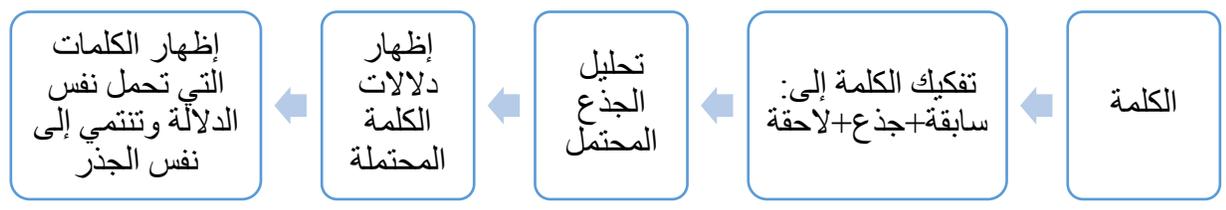
الشكل رقم (8): جدول تصريف الفعل الماضي السالم كَتَبَ.

## 6. المجال الدلالي:

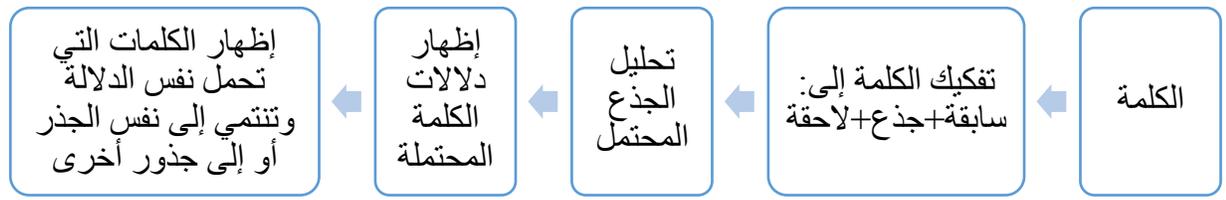
الأساس الذي تعمل عليه أداة التعرف الدلالي على الكلمة، مع الأخذ في الاعتبار أن الوحدة المعجمية الواحدة قد يحدد له أكثر من وحدة دلالية، وكذلك الوحدة الدلالية قد تستوعب عددًا من الوحدات المعجمية (المدخل) سواء أكانت تنتمي لنفس الجذر أو لجذور مختلفة، ولنا أن نتخيل جدوى هذا الأمر في عمليات البحث بالجذر ثم بالمعنى (مشتقات تنتمي لنفس الجذر تتشابه في المعنى)، أو البحث بالمرادفات (مشتقات تنتمي لجذور لغوية مختلفة تتشابه في المعنى)، ولمزيد من التوضيح نتعرف على آلية عمل أداة التعرف الدلالي.

### آلية عمل أداة التعرف الدلالي على الكلمة

ذكرنا أن المعجم يعمل وفق مستويين؛ صرفي ودلالي، ولا يمكن تصور أحد المستويين دون الآخر، وهذا ما يحقق الشمول الذي سعينا له من خلال التوليد الصرفي من ناحية، والدقة التي تتمثل في تقليص النتائج من خلال التعرف على معنى الكلمة من ناحية أخرى، حيث تعمل أداة التعرف الدلالي على الكلمة في تحديد المدخل والمشتقات المنتمية إلى نفس هذا المعنى، والشكلين التاليين يوضحان آلية عمل هذه الأداة، والتي تفيد في عمليتي البحث بالمعنى، والبحث بالمرادفات:



الشكل (9) آلية البحث بالمعنى



الشكل (10) آلية البحث بالمرادفات

### خيارات البحث وشكل/تكوين جملة البحث

يتضمن محرك البحث خيارات البحث اللغوية الآتية:

- البحث الافتراضي: هذا البحث يبحث بمشتقات كلمات البحث المُدخلة، أي بسوابقها لواحقها. وفيما يخص الأفعال فهو يبحث بالماضي والمضارع والأمر لكلمات البحث، بالإضافة إلى مشتقاتها.

- البحث بالأشكال الصرفية: هذا الخيار يبحث بجميع الأشكال الصرفية لكلمات البحث.
  - البحث بالمرادفات: هذا الخيار يبحث بجميع مرادفات كلمات البحث وأشكالها الصرفية.
- وتُشكّل كلمات البحث ما نسميه بجملة البحث؛ ونحن نستخدم نفس تكوين جملة البحث مُتّسقة مع محركات البحث العالمية:

- البحث يتم بكل كلمات البحث في النتائج (إلا في حالة استخدام العلاقة المنطقية أو).
- تُستخدم العلاقة المنطقية أو في جملة البحث بين كلمتين بحيث يمكن وجود أي منهما أو كلاهما في نتائج البحث.
- توضع كلمات البحث بين علامات التنصيص " " للبحث عنها كما هي.

### النتائج – أمثلة للبحث

ولاستشعار المميزات التي يتيحها محرك البحث، سوف نقوم بعرض النتائج الثلاث الأولى لكلمات البحث منظمة الدول المنتجة أو المصدرة للبترول في جزء من موقع الصحيفة العربية الأولى في الإمارات "الإمارات اليوم":

#### (أ) البحث الافتراضي:

كلمات البحث	منظمة	الدول	المنتجة	المصدرة	للبنترول
المعنى المحدد	المنظمات (مُنظّمة)	الدولة (دَوْل) / (دَوْل)	الإنتاج (مُنْتَج - مُنْتَج)	التصدير (مُصَدِّر - مُصَدِّر)	الوقود (البترول) (بَثْرُول/ بَثْرُول)

**النعيمة: ارتفاع أسعار النفط لا يُسعد السعودية - الإمارات اليوم**  
...وتتعاون مع أعضاء آخرين في منظمة الدول المصدرة للنفط أوبك على تحقيق هذا ..... نقل نفط إيران العضو في منظمة الدول المصدرة للنفط أوبك وقال مسؤولون تنفيذيون...

**قوات القذافي تتقدم باتجاه بنغازي وسط صمت دولي - الإمارات اليوم**  
...تتقدم باتجاه بنغازي وسط صمت دولي المصدر عواصم - وكالات التاريخ 17 ..... الاسلامي أمس أن وزراء خارجية دول المنظمة يعقدون بعد غد اجتماعاً استثنائياً...

**رفع أسعار البنزين في الدولة اعتباراً من الأربعاء المقبل - الإمارات اليوم**  
...المعروف في منطقة الخليج لأن الدول المنتجة للنفط والغاز يفترض أن تكون ..... الدولة اعتباراً من الأربعاء المقبل المصدر أحمد الشرييني دبي عبير عبدالحليم...

الشكل (11) نتائج البحث الافتراضي

(ب) البحث بالأشكال المصرفية:

كلمات البحث	منظمة	الدول	المنتجة	المصدرة	للبنترول
المعنى المحدد	المنظمات (مُنظَّمة)	الدولة (دَوْل) / دَوْل	الإنتاج (مُنْتَج - مُنْتِج)	التصدير (مُصَدِّر - مُصَدَّر)	الوقود (البنترول) (بَنْزُول / بَنْزُول)

**النعيمة: ارتفاع أسعار النفط لا يُسعد السعودية - الإمارات اليوم**  
...وتتعاون مع أعضاء آخرين في **منظمة الدول المصدرة** للنفط أوبك على تحقيق هذا ..... نقل نفط إيران العضو في **منظمة الدول المصدرة** للنفط أوبك وقال مسؤولون تنفيذيون...

**رفع أسعار البنزين في الدولة اعتباراً من الأربعاء المقبل - الإمارات اليوم**  
...باتريك كاستيلو قررت شركات توزيع **المنتجات البترولية** في **الدولة** رفع سعر لتر البنزين بأنواعه ..... المدير التنفيذي في شركة الإمارات **للمنتجات البترولية** إيبكو زيد عبدالرحمن الفقيدي أن...

**بدون عنوان**  
...والشي المحزن ان بلادنا هي **دولة منتجة** للنفط يعني المفروض يكون فيها ..... مب لازم يستمر في بيع **المنتجات البترولية** وممكن انه يشوف له شي...

الشكل (12) نتائج البحث بالأشكال المصرفية

(ج) البحث بالمرادفات:

كلمات البحث	منظمة	الدول	المنتجة	المصدرة	للبنترول
المعنى المحدد	المنظمات (مُنظَّمة)	الدولة (دَوْل) / دَوْل	الإنتاج (مُنْتَج - مُنْتِج)	التصدير (مُصَدِّر - مُصَدَّر)	الوقود (البنترول) (بَنْزُول / بَنْزُول)

**مادورو يستعد لوراثة السلطة بإشراف شافيز - الإمارات اليوم**  
...من النفط تقريباً وفقاً لتقارير **منظمة الدول المنتجة والمصدرة للنفط** أوبك فإن المشكلات الاقتصادية لفرنزويلا ..... يستعد لوراثة السلطة بإشراف شافيز **المصدر** ترجمة عوض خيري عن الإندبننت...

**النعيمة: ارتفاع أسعار النفط لا يُسعد السعودية - الإمارات اليوم**  
...وتتعاون مع أعضاء آخرين في **منظمة الدول المصدرة للنفط** أوبك على تحقيق هذا الهدف ..... نقل نفط إيران العضو في **منظمة الدول المصدرة للنفط** أوبك وقال مسؤولون تنفيذيون من...

**السعودية تزيد إنتاج النفط إلى 9 ملايين برميل يومياً - الإمارات اليوم**  
...الأحداث في ليبيا العضو في **منظمة الدول المصدرة للنفط** أوبك وتراجع إنتاجها من **النفط** إلى ذلك أفادت وكالة الطاقة ..... دون 111 دولاراً السعودية تزيد إنتاج **النفط** إلى 9 ملايين برميل يومياً...

الشكل (13) نتائج البحث بالمرادفات

## خاتمة

إن أرباب اللغة العربية من لغويين متخصصين أو تقنيين مهتمين بها، هم الأكثر تفهمًا لطبيعتها وأنظمتها الصرفية والنحوية والدلالية والصوتية، وهم الذين يدركون جيدًا مدى مرونة اللغة واستيعابها لكل ما هو جديد، وهم بذلك يستطيعون بإسهاماتهم في هذا المجال أن يعطوا اللغة إمكانات هائلة للارتقاء إلى العصر الرقمي، ويملكون البرهنة على قدرة اللغة الفاعلة في مختلف المعرفة والتبادل الثقافي من خلال استثمار الفرص الضخمة التي تتيحها شبكة المعلومات العالمية (الإنترنت).

ونأمل أن تسهم هذه التطبيقات التي نقدمها للغة العربية في مجال المعالجة الحاسوبية، في وضع خدمات جديدة ومفيدة للمستخدم العربي للكمبيوتر والإنترنت، وأن تعمل على إثراء المحتوى العربي الرقمي على الإنترنت، وأن تكون ثمرة تعاوننا كمتخصصين برمجيين ولغويين هي الارتقاء باللغة العربية والغوص في بحارها، واستخلاص بعض دُررها، ويكفيينا من ذلك أنها لغة أمتنا العربية ولغة دستورنا المقدس القرآن الكريم: ﴿كِتَابٌ فَصَّلْتُ آيَاتُهُ قُرْآنًا عَرَبِيًّا لِقَوْمٍ يَعْلَمُونَ﴾.



## الفهرس

- 1.....ملخص
- 2.....المقدمة
- 3.....مشكلة البحث بالكلمات العربية باستخدام محركات البحث التقليدية
- 3.....الخطوة الأولى: استيفاء شرط الشمول – استخدام القواعد الصرفية
- 4.....الخطوة الثانية: استيفاء شرط الدقة – استخدام خوارزميات للتعرف الدلالي على الكلمات
- 5.....مكونات النظام المتكامل
- 7.....التعريف بالمعجم الحاسوبي
- 7.....المدخل/الجذع
- 11.....آلية عمل أداة التعرف الدلالي على الكلمة
- 11.....خيارات البحث وشكل/تكوين جملة البحث
- 12.....النتائج – أمثلة للبحث
- 14.....خاتمة
- 15.....الفهرس